Genome Evolution of *Endomicrobia:* From Free-Living Bacteria to Intracellular Symbionts of Termite Gut Flagellates

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

am Fachbereich Biologie der Philipps-Universität Marburg

Universitätsstadt Marburg, 2015

vorgelegt von

Hao Zheng

aus Changchun, China

Die Untersuchungen zur folgenden Arbeit wurden von Oktober 2012 bis October 2015 am Max-Planck-Institut für Terrestrische Mikrobiologie in Marburg in der Forschungsgruppe von Prof. Dr. Andreas Brune durchgeführt.

Vom Fachbereich Biologie der Philipps-Universität Marburg als Dissertation angenommen am: 12.11.2015

Erstgutachter: Prof. Dr. Andreas Brune

Zweitgutachter: Prof. Dr. Uwe Maier

Tag der Disputation: 23.11.2015

Publications from this dissertation:

Zheng H, Dietrich C, Thompson CL, Meuser K, Brune A. 2015. Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microbes Environ*. 30:92–98.

Zheng H, Dietrich C, Radek R, Brune A. 2015. *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. *Environ Microbiol*. doi: 10.1111/1462-2920.12960.

Zheng H, Brune A. 2015. Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum *Elusimicrobia*). *Genome Announc*. 3:e00679-15.

Zheng H, Dietrich C, Hongoh Y, Brune A. Restriction-modification systems as mobile genetic elements in the evolution of an intracellular symbiont. *Mol Biol Evol.* doi: 10.1093/molbev/msv264

Manuscripts submitted or in preparation:

Zheng H, Dietrich C, Brune A. Loss and gain in *Endomicrobia* genomes: From free-living bacteria to intracellular symbionts (in preparation).

Contributions to manuscripts not included in this dissertation:

Zheng H, Herlemann DPR, Ikeda-Ohtsubo W, McHardy A, Lowry S, Goltsman E, Lapidus A, Rigoutsos I, Hugenholtz P, and A. Brune. Evidence for parallel evolution of '*Candidatus* Endomicrobium trichonymphae' in contrasting host environments (in preparation).

Thompson CL, Mikaelyan A, **Zheng H**, Rani P, Brune A. Multiple endosymbiotic events in the evolutionary history of *Endomicrobia* as symbionts of termite gut flagellates (in preparation).

Summary

Many eukaryotes harbor intracellular bacterial symbionts that are believed to confer beneficial traits to their hosts. Several bacterial lineages in the class of *Endomicrobia* are frequently encountered as intracellular symbionts of termite gut flagellates. They represent a deep-branching lineage in the phylum of *Elusimicrobia* that comprise both endosymbiotic and putatively free-living lineages. Since the acquisition of the endosymbionts probably occurred around 40–70 million years ago, the genomes of the endosymbionts are still in an early stage of genome reduction. Therefore, genomic information on free-living members of this class would provide a perfect model to investigate the mechanisms of genome evolution that have contributed to the evolution of such a young endosymbiosis.

In my dissertation, I first describe the endosymbiotic population of *Endomicrobia* in single flagellate hosts with highly resolved analyses. These endosymbionts are strictly vertically transmitted by their respective hosts indicating a frequent population bottleneck during transmission for *Endomicrobia*. Moreover, I report the isolatation of *Endomicrobium proavitum*, the first cultured representative of the class *Endomicrobia* and a close relative of "*Candidatus* Endomicrobium trichonymphae", the intracellular symbionts that colonize the flagellates of the same termite species. This free-living, obligately anaerobic ultramicrobacterium has an unusual cell cycle and fixes nitrogen employing a set of *nif* genes (Group IV) that so far had been considered not to encode a functional nitrogenase. Its circular genome (1.59 Mbp) is substantially larger than that of "*Ca.* E. trichonymphae" strain Rs-D17, and many of the pathways that are disrupted by pseudogenes in the endosymbiont are intact in *E. proavitum*. However, during the process of symbiosis, the endosymbionts also appear to have

acquired novel pathways to alter their fermentation strategy, which have apparently helped the establishment of the associations.

The comparative analyses not only revealed that the genome of strain Rs-D17 shows a typical purifying selection throughout the genome independent on the gene functions but also provided evidence for massive genome rearrangements in the endosymbionts. Surprisingly, the rearrangement sites are typically flanked by restriction-modification genes that are abnormally dense in the genome of strain Rs-D17 but entirely absent from its free-living relative. This is the first case where RM systems acting as mobile genetic elements are responsible for shaping a bacterial genome in the early stages of endosymbiosis.

Zusammenfassung

Viele Eukaryoten besitzen intrazelluläre bakterielle Symbionten, die ihren Wirt vermutlich funktionell komplementieren. Die Klasse *Endomicrobia* , eine tief abzweigende Gruppe von Bakterien, wurden in der Literatur zunächst als intrazelluläre Symbionten von Termitendarm-Flagellaten beschrieben, beinhalten jedoch auch frei im Darm lebende Vertreter. Da die Aufnahme von Endomicrobia vor 40–70 Millionen Jahren stattfand, befinden sich die entsprechenden Genome noch in einer frühen Phase der Genomreduktion.

Innerhalb dieser Dissertation beschreibe ich endosymbiontische *Endomicrobia*-Populationen innerhalb verschiedener individueller Flagellaten-Wirte mit hochauflösenden Methoden. Die Ergebnisse zeigen deutlich, dass intrazelluläre *Endomicrobia* vertikal von ihrem Wirt übertragen werden und damit langfristig einer genetischen Verarmung unterliegen.

Weiterhin beschreibe ich die Isolation von *Endomicrobium proavitum*, dem ersten kultivierten Vertreter der Klasse *Endomicrobia*, der nah verwandt mit dem intrazellulären Termitendarm-Flagellaten-Symbiont "*Candidatus* Endomicrobium trichonymphae" ist. *Endomicrobium proavitum* ist ein freilebendes Ultramikrobakterium, das einen ungewöhnlichen Zellzyklus aufweist und molekularen Stickstoff mit einer Gruppe-4-Nitrogenase fixieren kann, die bisher als funktionsunfähig galt. Im Vergleich zu "*Ca.* E. trichonymphae" besitzt *Endomicrobium proavitum* ein deutlich größeres Genom (1,59 Mbp gegenüber 1,13 Mbp) und zeigt eine wesentlich geringere Anzahl der in "*Ca.* E. trichonymphae" häufigen Pseudogene, die viele biochemisch relevante Prozesse außer Kraft setzen. Im Gegenzug scheint "*Ca.* E. trichonymphae" neue Stoffwechselwege erworben zu haben um seinen fermentativen Stoffwechsel

verändern zu können, was offensichtlich hilfreich für die Assoziation mit dem Flagellaten-Wirt war.

Vergleichende Genomik zeigte, dass der intrazelluläre Symbiont "*Ca.* E. trichonymphae" unter starkem negativem Selektionsdruck steht, der auf alle Gene unabhängig von der Funktion der codierten Enzyme wirkt. Weiterhin habe ich herausgefunden, dass in "*Ca.* E. trichonymphae" das Genom häufig reorganisiert wurde. Regionen eindeutiger genomischer Reorganisation sind häufig von Restriktions-Modifikations(RM)-Systemen umgeben, die ungewöhnlich häufig in "*Ca.* E. trichonymphae" zu finden sind, aber im freilebenden Verwandten *Endomicrobium proavitum* nicht vorkommen. Damit ist dies ist der erste Beleg für eine durch RM-Systeme verursachte Genom-Reorganisation im jungen Stadium einer Endosymbiose.

Contents

Chapter 1

General Introduction

Termite and the symbiotic flagellates

It is now generally accepted that termites are social insects that evolved from cockroaches (Bourguignon et al. 2015), and the first traces of termites date back to the Jurassic Period, about 195 million years ago (Mya) (Xing et al. 2013). An important transition point of the evolution is the acquisition of cellulolytic gut flagellates, which provide an essential contribution to the lignocellulose digestion, by a common ancestor of termites and the wood-feeding cockroaches (Cryptocercidae), However, in the members of the youngest termite family (Termitidae, also called “higher termites”), which arose around 50 Mya (Bourguignon et al. 2015), the gut is no longer associated with cellulolytic flagellates (Ohkuma and Brune 2011).

The termite gut microbiota comprises all three domains of life: Bacteria, Archaea, and Eukarya (flagellates are only dominating the lower termite guts). The termite gut flagellates contain diverse lineages in the phyla *Parabasalia* and *Preaxostyla* (order *Oxymonadida*) that are mostly unique to lower termites (Brune 2014). Usually, the flagellates are associated with large numbers of prokaryotic symbionts that occur both on the cell surface, in the cytoplasm (Hongoh and Ohkuma 2010), and in some special cases, even inside the nucleus of the hosts (Sato et al. 2014). It has been suggested that the flagellates rely on the prokaryotic symbionts for the nitrogen metabolism and other important nutritional provisions (Brune 2011; Hongoh 2011).

Endomicrobia – Discovery, localization and morphology

Endomicrobia are a specific lineage of symbionts that were found in large populations in the termite gut. Initially, they were discovered by a molecular study

of bacterial diversity in *Reticulitermes speratus* comprising a large portion of the 16S rRNA gene sequences showing high divergence to any other bacterial sequences back then (Ohkuma and Kudo 1996). They were named Termite Group 1 (TG1) and, due to their deep-branching phylogeny, referred to as a candidate phylum (Hugenholtz et al. 1998). Although the intracellular localization of a TG1 phylotype in the parabasalid flagellate *Trichonympha agilis* was reported at several conferences (Eldridge et al. 1997; Ohkuma et al. 2001), it took several years until their status as endosymbionts of both parabasalid and oxymonadid flagellates in *R. speratus* and other species of lower ternites was documented in the literature (Stingl et al. 2005; Ikeda-Ohtsubo et al. 2007; Ohkuma et al., 2007). There abundant presence also in many other gut flagellates of lower termites (Table 1.1) explained why members of the TG1 phylum make up a substantial portion of the gut community among members of the of the termite genus *Reticulitermes* (Hongoh et al. 2003; Hongoh et al. 2005; Yang et al. 2005) and other species of lower termites (Dietrich et al. 2014). Interestingly, TG1 members were also detected from higher termites, a cockroach, and the bovine rumen that are typically lack of gut flagellates, which form monophyletic cluster in the phylogenetic tree (Ohkuma et al. 2007; Ikeda-Ohtsubo et al. 2010).

Transmission electron microscopy of ultrathin sections of TG1 endosymbionts revealed that they are small rods (approximately 0.6 × 0.3 µm). They show an unusual spindle-shaped appearance with distinctly tapered cell poles and are surrounded by two membranes. In the case of *Trichonympha* endosymbionts, they formed tubular extensions (0.5 µm long) extending into the cytoplasm (Stingl et al. 2005). The endosymbionts seem to be evenly distributed in the cytoplasm of their hosts and sometimes aggregate in certain portion of the host cells (Stingl et al. 2005; Ohkuma et al. 2007). The population size of TG1

Table 1.1. Flagellates possess *Endomicrobia* as endosymbionts determined by picked flagellate cells.

Flagellate	Termite	Reference
Trichonymphida		
Eucomonympha sp.	*Hodotermopsis sjoestedti*	Ikeda-Ohtsubo et al., 2007
Trichonympha agilis	*Reticulitermes flavipe* *Reticulitermes Hesperus* *Reticulitermes lucifugus* *Reticulitermes santonensis* *Reticulitermes speratus* *Reticulitermes* sp. OR1	Stingl et al., 2005; Ohkuma et al., 2007; Ikeda-Ohtsubo et al., 2007; Zheng et al., 2015
Trichonympha campanula	*Zootermopsis nevadensis*	Zheng et al., 2015
Trichonympha collaris	*Zootermopsis nevadensis*	Ikeda-Ohtsubo et al., 2007; Ikeda-Ohtsubo and Brune 2009; Zheng et al., 2015
Trichonympha postcylindrica	*Zootermopsis nevadensis*	Zheng et al., 2015
Trichonympha sphaerica	*Zootermopsis nevadensis*	Ikeda-Ohtsubo and Brune 2009; Zheng et al., 2015
Trichonympha spp.	*Hodotermopsis sjoestedti*	Ohkuma et al., 2007; Ikeda-Ohtsubo and Brune 2009
Oxymonadida		
Dinenympha exilis	*Reticulitermes speratus*	Ohkuma et al., 2007
Dinenympha rugosa	*Reticulitermes speratus*	Ohkuma et al., 2007
Dinenympha sp.	*Hodotermopsis sjoestedti*	Ikeda-Ohtsubo et al., 2007
Oxymonas sp.	*Neotermes castaneus*	Ikeda-Ohtsubo et al., 2007
Pyrsonympha grandis	*Reticulitermes speratus*	Ohkuma et al., 2007
Pyrsonympha vertens	*Reticulitermes speratus*	Stingl et al., 2005
Cristamonadida		
Calonympha sp.	*Neotermes castaneus*	Ikeda-Ohtsubo et al., 2007
Deltotrichonympha sp.	*Mastotermes darwiniensis*	Ikeda-Ohtsubo et al., 2007
Devescovina lemniscata	*Cryptotermes domesticus*	Ohkuma et al., 2007
Devescovina sp.	*Neotermes castaneus*	Ikeda-Ohtsubo et al., 2007
Gigantomonas herculea	*Hodotermes mossambicus*	Ohkuma et al., 2007
Macrotrichomonas sp.	*Glyptotermes fuscus*	Ohkuma et al., 2007
Metadevescovina cuspidata	*Incisitermes minor*	Ohkuma et al., 2007
Joenia annectens	*Kalotermes flavicollis*	Ohkuma et al., 2007
Joenoides intermedia	*Hodotermes mossambicus*	Ohkuma et al., 2007

bacteria in a single flagellate cell was estimated up to about 4,000 (Ohkuma et al. 2007). Since the flagellates densely colonize the hindgut in all lower termites, their endosymbiotic TG1 bacteria contribute a large proportion of the gut community (Stingl et al. 2005; Ohkuma et al. 2007). TG1 phylotypes make up a large proportion of the bacterial 16S rRNA clone library, up to 10% (Hongoh et al. 2003; Hongoh et al. 2005), which might be even underestimated considering they contain only one single copy of the *rrn* operon (Hongoh et al. 2008).

Stingl et al. (Stingl et al. 2005) suggested the candidate phylum 'Endomicrobia' for the deep-branching lineage of TG1 symbionts, comprising exclusively of clones from the guts of lower termites and wood-feeding cockroaches of the genus *Cryptocercus*. The subsequent screening for 16S rRNA gene sequences of TG1 bacteria revealed more uncultured bacteria from diverse environments that fell within the Endomicrobia clade (Herlemann et al. 2007). Later, the first cultivated representative, *Elusimicrobium minutum*, a strictly anaerobic ultramicrobacterium, was isolated from filter-sterilized hindgut fluid of a scarab beetle larva and belongs. It falls into a different, deep-branching lineage of the whole clade. It was then proposed *Elusimicrobia* for the entire phylum and *Endomicrobia* was demoted to the level of a taxonomic class (Geissinger et al. 2009).

Genomes and functional roles

While it is important to clarify the functions of individual species in termite gut microbiota, especially the dominant and termite-specific lineages, most of them still remain uncultivated (Brune 2006). Hongoh et al. (2008) obtained the complete genome of '*Ca*. Endomicrobium trichonymphae' strain Rs-D17, the endosymbiont of a single cell of *T. agilis*, from limited number of cells. The

genome is very small (1.13 Mbp) and many pathways have been collapsed with lots of pseudogenes, which is typical for intracellular symbionts. However, strain Rs-D17 retained the capacity of 15 amino acids and many cofactors biosynthesis, which are critically limiting in wood diet, indicating its potential role in the nutrition provision to the flagellate hosts. Such a symbiotic relationship might drive their evolutionary directions, such as cospeciation (Noda et al. 2007; Desai et al. 2010).

Herlemann et al. (Herlemann et al. 2009) determined the genome of the distantly related isolate, *Elusimicrobium minutum*. It is slightly larger (1.64 Mbp) and pseudogenes are almost absent. However it is quite distant to the genome of strain Rs-D17, the evolutionary origin of the *Endomicrobia* lineage is still obscure. There is a particular need for genomic information of close and free-living relatives to better understand the genome reduction process of this intracellular symbiosis association.

Vertical transmission and acquisition history

It was noticed that the phylotypes of *Endomicrobia* from the same flagellate species form a monophyletic cluster (Ikeda-Ohtsubo et al., 2007; Ohkuma et al. 2007). Ikeda-Ohtsubo and Brune (Ikeda-Ohtsubo and Brune 2009) documented the congruence of the phylogenies of the TG1 symbionts and their host flagellates in the genus *Trichonympha*, which suggests that the endosymbionts might be propagated by vertical transmission in this host lineage. However, due to the relatively small number of sequences in the libraries and the fact that they were obtained from pools of flagellate cells, it could not be excluded that more than one phylotype of *Endomicrobia* is present within each *Trichonympha* cell.

While the monophyly of the whole *Endomicrobia* clade is well supported, the origin of the insect lineage remains unclear, particularly because the lack of phylotypes from other environments and presence of the next clade from the bovine rumen (Ohkuma et al. 2007). Endomicrobia are present also in cockroaches indicating that endosymbiotic *Endomicrobia* were derived from free-living relatives that arrived before the separation of cockroach and termite line dated back to 130 Mya (Ohkuma et al. 2009). Although it was shown that a certain group of *Trichonympha* harbor *Endomicrobia* symbionts, their absence in the basal clades of the genus (Clusters II and III) postpones the establishment of endosymbiosis between *Endomicrobia* and the common ancestor of *Trichonympha* Cluster I to about 40–70 Mya, long after the flagellates had been acquired by the ancestor of termites and *Cryptocercus* cockroaches (Ikeda-Ohtsubo and Brune 2009). Novel phylotypes of Endomicrobia are also detected after elimination of the gut flagellates along with their abundant symbionts, which indicates that putatively free-living forms of Endomicrobia are present also in lower termites (Ikeda-Ohtsubo et al. 2010). This co-occurrence of free-living and endosymbiotic Endomicrobia makes an attractive model for the analysis of the mechanism of endosymbiosis and the genome evolution during this process.

Aims of this thesis

Population structure. Although Endomicrobia show congruent phylogeny with their flagellate hosts, it remains unclear whether each host species harbor the same symbionts and whether a single flagellate cell possesses more than one *Endomicrobia* endosymbionts phylotype. I aim to clarify the situation by characterizing the *Endomicrobia* populations of manipulated single flagellate cells,

particularly using the internal transcribed spacer (ITS) region between the rRNA genes to resolve even low levels of strain variations.

Isolation. Thus far, the only cultivated representative of the *Elusimicrobia* phylum is *Elusimicrobium minutum*, which belongs to a different, deep-branching lineage of the phylum. The entire *Endomicrobia* clade still remains uncultured. Inspired by the observation of abundant putatively free-living lineages in lower termites upon starch-feeding, I try to enrich free-living *Endomicrobia* strains from filter-sterilized gut homogenates of starch-fed lower termites and isolate the first representative of *Endomicrobia*.

Genome evolution. The genome sequence of strain Rs-D17 provides first insights into the ongoing reductive evolution of the endosymbionts. The co-occurrence of both endosymbiotic and free-living relatives in the termite gut promises a particularly good example for studying the evolution of an intracellular symbiosis. I will obtain the complete genome sequence of the free-living *Endomicrobia* and the detailed comparative analysis will allow us to understand the mechanism of the genome evolution during this relative early associated endosymbiotic association.

References

Bourguignon T, Lo N, Cameron SL, Sobotnik J, Hayashi Y, Shigenobu S, Watanabe D, Roisin Y, Miura T, Evans TA. 2015. The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Mol Biol Evol*. 32:406–421.

Brune A. 2006. Symbiotic associations between termites and prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH ,Stackebrandt E, editors.Prokaryotes: A Handbook on the Biology of Bacteria. New York: Springer, pp. 439–474.

Brune A. 2011. Microbial symbioses in the digestive tract of lower termites. In: Rosenberg E ,Gophna U, editors.Beneficial Microorganisms in Multicellular Life Forms. Heidelberg: Springer, pp. 3–25.

Brune A. 2014. Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol*. 12:168–180.

Desai MS, Strassert JFH, Meuser K, Hertel H, Ikeda-Ohtsubo W, Radek R, Brune A. 2010. Strict cospeciation of devescovinid flagellates and *Bacteroidales* ectosymbionts in the gut of dry-wood termites (Kalotermitidae). *Environ Microbiol*. 12:2120–2132.

Eldridge ML, Goss S, Gunderson J. 1997. Identification of a prokaryotic symbiont of the termite flagellate *Trichonympha*. In *The 97th Annual Meeting of the American Society for Microbiology*. Miami Beach, USA, pp. N-037.

Geissinger O, Herlemann DPR, Möschel E, Maier UG, Brune A. 2009. The ultramicrobacterium "*Elusimicrobium minutum*" gen. nov., sp. nov., the first cultivated representative of the Termite Group 1 phylum. *Appl Environ Microbiol*. 75:2831–2840.

Herlemann DPR, Geissinger O, Brune A. 2007. The termite group I phylum is highly diverse and widespread in the environment. *Applied and Environmental Microbiology*. 73:6682–6685.

Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P, Brune A. 2009. Genomic analysis of "*Elusimicrobium minutum*," the first cultivated representative of the phylum "*Elusimicrobia*" (formerly Termite Group 1). *Appl Environ Microbiol*. 75:2841–2849.

Hongoh Y, Ohkuma M, Kudo T. 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). *FEMS Microbiol Ecol*. 44:231–242.

Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N, Kudol T. 2005. Intra- and interspecific comparisons of bacterial diversity and community structure support

coevolution of gut microbiota and termite host. *Appl Environ Microbiol*. 71:6590–6599.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci U S A*. 105:5555–5560.

Hongoh Y, Ohkuma M. 2010. Termite gut flagellates and their methanogenic and eubacterial symbionts. In: Hackstein JHP, editors.(Endo)Symbiotic Methanogenic Archaea. Heidelberg: Springer, pp. 55–79.

Hongoh Y. 2011. Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell Mol Life Sci*. 68:1311–1325.

Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol*. 180:4765–4774.

Ikeda-Ohtsubo W, Desai M, Stingl U, Brune A. 2007. Phylogenetic diversity of 'Endomicrobia' and their specific affiliation with termite gut flagellates. *Microbiology*. 153:3458–3465.

Ikeda-Ohtsubo W, Brune A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol Ecol*. 18:332–342.

Ikeda-Ohtsubo W, Faivre N, Brune A. 2010. Putatively free-living '*Endomicrobia*' – ancestors of the intracellular symbionts of termite gut flagellates? *Environ Microbiol Rep*. 2:554–559.

Noda S, Kitade O, Inoue T, Kawai M, Kanuka M, Hiroshima K, Hongoh Y, Constantino R, Uys V, Zhong J, et al. 2007. Cospeciation in the triplex symbiosis of termite gut protists (Pseudotrichonympha spp.), their hosts, and their bacterial endosymbionts. *Molecular Ecology*. 16:1257–1266.

Ohkuma M, Kudo T. 1996. Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. *Appl Environ Microbiol*. 62:461–468.

Ohkuma M, Noda S, Iida T, Kudo T. 2001. Phylogenetic identification of endosymbionts of the flagellated protists in the gut of termites. In *The 9th International Symposium on Microbial Ecology (ISME-9).* Amsterdam, The Netherlands, pp. 280.

Ohkuma M, Sato T, Noda S, Ui S, Kudo T, Hongoh Y. 2007. The candidate phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol Ecol*. 60:467–476.

Ohkuma M, Noda S, Hongoh Y, Nalepa CA, Inoue T. 2009. Inheritance and diversification of symbiotic trichonymphid flagellates from a common ancestor of termites and the cockroach *Cryptocercus*. *Proc R Soc B*. 276:239–245.

Ohkuma M, Brune A. 2011. Diversity, structure, and evolution of the termite gut microbial community. In: Bignell DE, Roisin Y ,Lo N, editors.Biology of Termites: A Modern Synthesis. Dordrecht: Springer, pp. 413–438.

Sato T, Kuwahara H, Fujita K, Noda S, Kihara K, Yamada A, Ohkuma M, Hongoh Y. 2014. Intranuclear verrucomicrobial symbionts and evidence of lateral gene transfer to the host protist in the termite gut. *ISME J*. 8:1008–1019.

Stingl U, Radek R, Yang H, Brune A. 2005. "*Endomicrobia*": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol*. 71:1473–1479.

Xing LD, Roberts EM, Harris JD, Gingras MK, Ran H, Zhang JP, Xu X, Burns ME, Dong ZM. 2013. Novel insect traces on a dinosaur skeleton from the Lower Jurassic Lufeng Formation of China. *Palaeogeogr Palaeocl*. 388:58–68.

Yang H, Schmitt-Wagner D, Stingl U, Brune A. 2005. Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ Microbiol*. 7:916–932.

Zheng H, Dietrich C, Thompson CL, Meuser K, Brune A. 2015. Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microbes Environ*. 30:92–98.

Chapter 2

Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.)

Published in *Microbes and Environments* (2015) 30:92–98.

Hao Zheng[1], Carsten Dietrich[1], Claire L. Thompson[2], Katja Meuser[1], Andreas Brune[1,2]

[1]Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany; [2]LOEWE Center for Synthetic Microbiology (SYNMIKRO), Philipps-Universität Marburg, Marburg, Germany

Contributions: H.Z. planned and carried out experiments, analyzed data, discussed results, and wrote the manuscript. C.D. analyzed data, discussed results, and contributed to the manuscript. C.L.T. and K.M. carried out experiments. A.B. conceived the study, discussed results, wrote the manuscript, and secured funding.

Abstract

The gut microbiota of many phylogenetically lower termites is dominated by cellulolytic flagellates of the genus *Trichonympha*, which are consistently associated with bacterial symbionts. In the case of "Endomicrobia", an unusual lineage of endosymbionts of the *Elusimicrobia* phylum that is present also in other gut flagellates, previous studies have documented strict host specificity, leading to cospeciation of "*Candidatus* Endomicrobium trichonymphae" with their respective flagellate hosts. However, it remained unclear if one *Trichonympha* species can harbor more than one Endomicrobia phylotype. In this study, we picked single *Trichonympha* cells from the gut of *Zootermopsis nevadensis* and *Reticulitermes santonensis* and characterized their Endomicrobia populations based on the internal transcribed spacer (ITS) region sequences. We found that each host cell harbors a homogeneous population of symbionts that were specific for their respective host species but phylogenetically distinct between each host lineage, corroborating that cospeciation is caused by vertical inheritance. The experimental design of our study also allowed identification of an unexpectedly large amount of tag-switching between samples, which indicates that any high-resolution analysis of microbial community structure using the pyrosequencing technique has to be interpreted with great caution.

Introduction

The termite hindgut is colonized by dense assemblages of prokaryotes (bacteria and archaea) and eukaryotic protists that play essential roles in both digestion and host nutrition (Brune and Ohkuma 2011; Hongoh 2011). The gut microbiota of the phylogenetically lower termites is dominated by cellulolytic flagellates that are unique to termites and comprise diverse lineages from the phyla *Parabasalia* and *Preaxostyla* (order *Oxymonadida*) (Brune 2014). In most cases, the flagellates are associated with large numbers of prokaryotic symbionts that colonize both the surface and cytoplasm (Hongoh and Ohkuma 2010; Ohkuma and Brune 2011) and sometimes even the nucleus of their hosts (Sato et al. 2014). There is strong evidence that the symbionts complement the nitrogen metabolism of the flagellates and have an important nutritional role in the hindgut microecosystem (Hongoh and Ohkuma 2010; Brune 2011; Hongoh 2011; Brune 2014).

The predominant flagellates in the hindgut of several termite lineages are hypermastigid flagellates of the genus *Trichonympha*. Originally discovered in *Reticulitermes* spp. (Rhinotermitidae) (Leidy 1881; Grassi and Sandias 1897; Koidzumi 1921), members of the genus have been described also in several other termite families (Termopsidae, Kalotermitidae) and in their sister group, the wood-feeding cockroaches of the genus *Cryptocercus* (for references, see Yamin 1979). While some termite species contain only a single species of *Trichonympha*, others harbor several representatives of this genus.

Termites of the genus *Zootermopsis* harbor at least four *Trichonympha* species. *Trichonympha campanula* (Kofoid and Swezy 1919a), *Trichonympha sphaerica* (Kofoid and Swezy 1919b), and *Trichonympha collaris* (Kirby 1932)

have long been recognized based on differences in morphology; the fourth species was discovered using molecular approaches. The first evidence of the existence of this fourth species was the presence of an additional phylotype of small subunit (SSU) rRNA gene sequences in a picked cell suspension of *T. campanula* from *Zootermopsis nevadensis* that was differentiated from the cells of *T. campanula* by fluorescence *in situ* hybridization (Ikeda-Ohtsubo 2007). A few years later, this species, *Trichonympha postcylindrica,* was described based on specimens from the gut of *Zootermopsis angusticollis* (Tai et al. 2013).

Although the same four *Trichonympha* species seem to be present in both *Z. nevadensis* and *Z. angusticollis* (Kirby 1932; Ikeda-Ohtsubo 2007; Tai et al. 2013), their relative abundance may differ even among individuals from the same colony. Moreover, there is a considerable range of sequence divergence within the populations of each *Trichonympha* species in *Z. angusticollis* (Tai et al. 2013). The current knowledge about the diversity of *Trichonympha* species in the two *Zootermopsis* species, including the hitherto unpublished sequence of *T. postcylindrica* from *Z. nevadensis* (Ikeda-Ohtsubo 2007), is summarized in Fig. 2.1.

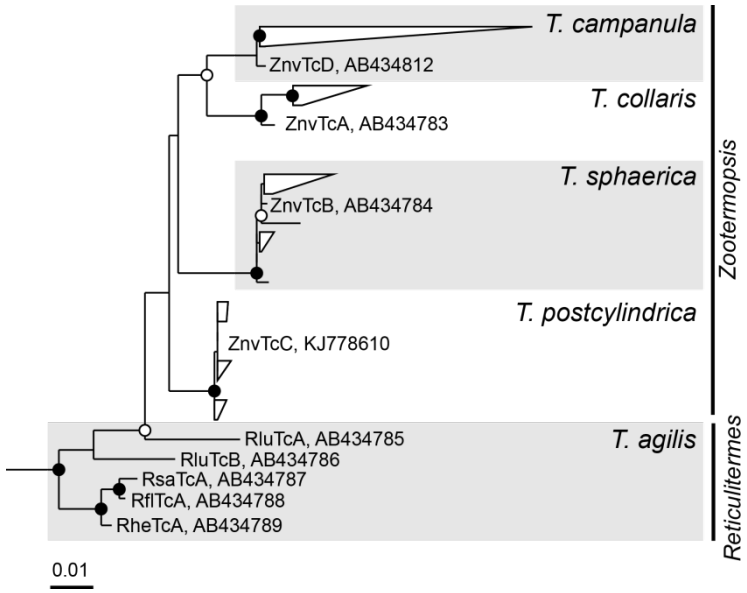

Fig. 2.1. Diversity of the 18S rRNA gene sequences of *Trichonympha* flagellates in *Zootermopsis nevadensis*, *Zootermopsis angusticollis* (grouped nodes), and different *Reticulitermes* species (Rlu, *Reticulitermes lucifugus*; Rsa, *Reticulitermes santonensis*; Rfl, *Reticulitermes flavipes*; Rhe, *Reticulitermes hesperus*). The maximum-likelihood tree was rooted with *Trichonympha* species of Clusters II and III. Circles indicate node support (○, > 80%; ●, > 95% bootstrap confidence).

Members of the genus *Trichonympha* fall into three distinct phylogenetic clusters (Carpenter et al. 2009; Ikeda-Ohtsubo and Brune 2009; Ohkuma et al. 2009), and all are associated with bacterial symbionts, quite often more than one bacterial species per flagellate host (Sato et al. 2009; Strassert et al. 2012; Sato et al. 2014). All members of *Trichonympha* Cluster I harbor large populations of a specific intracellular symbiont, "*Candidatus* Endomicrobium trichonymphae" (Stingl et al. 2005; Ikeda-Ohtsubo et al. 2007; Ohkuma et al. 2007), which belongs to a distinct lineage of uncultured, insect-associated bacteria ("Endomicrobia") in the *Elusimicrobia* phylum (Brune 2012). The presence of homogeneous bacterial symbiont populations in a single flagellate host cell (Hongoh et al. 2008) and an almost perfect congruence of the phylogenies of the

bacterial symbionts and their host flagellates (Ikeda-Ohtsubo and Brune 2009) suggest that "*Ca.* Endomicrobium trichonymphae" is propagated exclusively by vertical transmission (cytoplasmic inheritance) in the *Trichonympha* lineage. In this case, one would expect the same number of Endomicrobia species as *Trichonympha* species in a given termite. This is in apparent contradiction with the composition of a metagenomic library constructed from DNA prepared from the enrichment of "*Ca.* Endomicrobium trichonymphae" in the gut of *Z. nevadensis*, which contains a larger number of 16S rRNA phylotypes of Endomicrobia than host species (IMG Project ID: Gi01566) (Ikeda-Ohtsubo 2007). It remains unclear whether all individuals of each host species harbor the same symbionts and whether a single *Trichonympha* cell carries more than one Endomicrobia phylotype.

To clarify the situation, we isolated single *Trichonympha* cells from the gut of *Z. nevadensis* and *R. santonensis* and characterized their Endomicrobia populations by high-throughput sequencing, using the internal transcribed spacer (ITS) region between their rRNA genes to resolve even low levels of strain variation within both symbionts and hosts.

Results

Isolation and identification of single *Trichonympha* cells

We successfully amplified the ITS regions of host flagellates (Fig. 2.2A) and Endomicrobia symbionts (Fig. 2.2B) from the whole-genome amplification products obtained from picked Trichonympha cells from *Z. nevadensis* (34 of 60 cells) and R. santonensis (9 of 20 cells). Direct sequencing of the PCR products obtained with flagellate-specific primers yielded clean signals for all species

except for *T. campanula*, where multiple bases at several positions of the trace file indicated sequence polymorphism in the ITS region (Fig. S2.1).

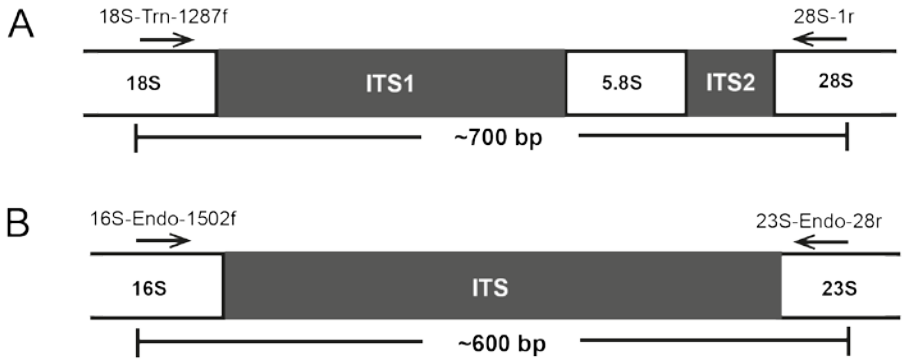

Fig. 2.2. Strategy used for the PCR amplification of the ITS regions of (A) *Trichonympha* flagellates and (B) their Endomicrobia symbionts. The primers, their target positions in the proximal regions of the flanking rRNA genes, and the resulting amplicon length are indicated.

Comparative sequence analysis showed that each of the sequences obtained from the morphotypes of *T. sphaerica*, *T. postcylindrica*, *T. collaris*, and *T. campanula* in *Z. nevadensis* clustered with their respective relatives from *Z. angusticollis* (Fig. 2.3), confirming the morphological assignment of the picked flagellates. The nine sequences from the morphotype of *T. agilis* formed a sister group of the Trichonympha species from *Z. nevadensis*.

Endomicrobia populations in single *Trichonympha* cells

Pyrosequencing analysis of PCR products obtained with Endomicrobia-specific primers yielded variable read numbers per flagellate sample (after quality control, see Fig. S2.2). Although the number of Endomicrobia phylotypes (99% sequence identity) obtained from all samples (17 phylotypes from 45 single host cells) was much larger than the number of flagellate species investigated (5

species), they all clustered more or less closely according to their respective hosts (Fig. 2.4A).

While the Endomicrobia obtained from nine different cells of *T. agilis* comprised only two closely related phylotypes, the symbionts of the other flagellates, particularly *T. postcylindrica* and *T. campanula*, were far more diverse (Fig. 2.4A). Phylogenetically distinct flagellates carried different phylotypes of symbionts, and even the Endomicrobia in host cells with identical ITS sequences sometimes differed slightly in their respective phylotype (e.g., *T. sphaerica* ZnvTrn10 and ZnvTrn48).

When we analyzed the diversity of Endomicrobia sequences obtained from whole-gut DNA of *Z. nevadensis*, we found several additional phylotypes (represented by black branches in Fig. 2.4A) within the radiation of the Endomicrobia sequences retrieved from picked *Trichonympha* cells, which indicated that the *Trichonympha* populations in this termite were not exhaustively sampled. Nonetheless, the vast majority of the Endomicrobia in *Z. nevadensis* consisted of phylotypes retrieved from *T. collaris* (41%) and *T. sphaerica* (33%), while the diverse populations of Endomicrobia associated with *T. postcylindrica* (13%) and *T. campanula* (6%) represented only a small part of the community (Fig. 2.4B). In the case of *R. santonensis*, the two phylotypes of Endomicrobia from *T. agilis* accounted for 23% of the phylotypes in the gut homogenate, and the majority of the reads clustered with ITS sequences of picked *Pyrsonympha* flagellates (unpublished results), which form a distinct branch in the phylogeny of Endomicrobia (Stingl et al. 2005).

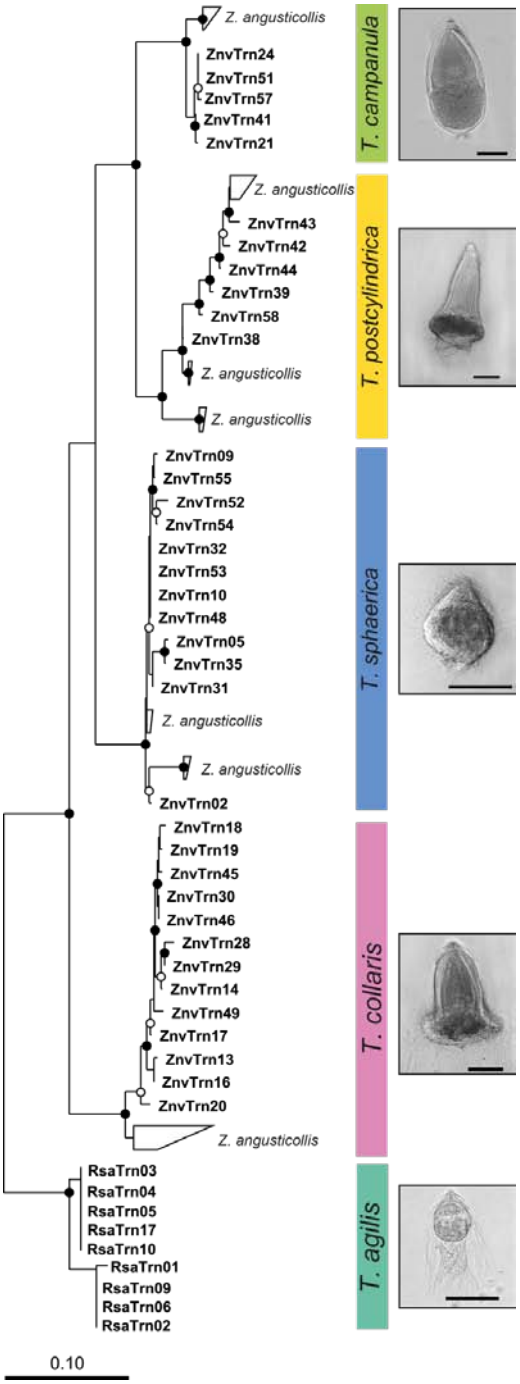

Fig. 2.3 Neighbor-joining tree of the ITS sequences obtained from picked flagellates, illustrating the phylogenetic diversity of the *Trichonympha* species from *Zootermopsis nevadensis* (this study) and their relationship to those from *Z. angusticollis* (sequences are grouped). The tree is based on an alignment of 450 base positions and was rooted with the ITS sequences of *T. agilis* from *R. santonensis*. Phase-contrast photomicrographs illustrate the morphology of the respective species (scale bars = 50 µm). Circles indicate node support (○, > 80%; ●, >95% bootstrap confidence).

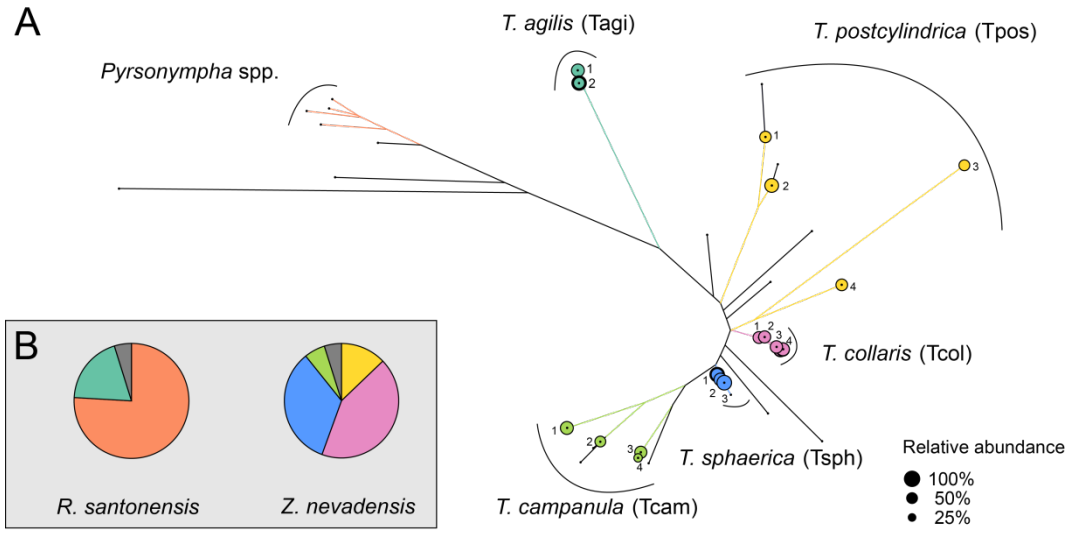

Fig. 2.4. (A) Maximum-likelihood tree of all Endomicrobia phylotypes (99% sequence similarity) detected in the whole-genome amplification (MDA) products of all single flagellate cells. Branches shared by phylotypes from the same host species are color coded; the area of the circles indicates the relative abundance of the major phylotype in each library. Black dots indicate the position of phylotypes that originated from whole-gut samples of *Z. nevadensis* or *R. santonensis*. (B) Pie charts indicate the relative abundance of the different Endomicrobia clusters in the whole gut samples. The colors represent the major phylotypes from picked *Trichonympha* cells; phylotypes from whole-gut samples are shown in grey. Phylotypes in the *Pyrsonympha* cluster were identified based on the sequences retrieved from picked flagellates from *R. santonensis* (unpublished results).

Pyrosequencing artifacts

Although each of the pyrotag libraries of Endomicrobia from single flagellates consistently contained only one major phylotype (Fig. 2.5), they always comprised a smaller fraction of reads that were identical with the dominant phylotypes in other libraries. The shared presence of the same Endomicrobia phylotype in different flagellate samples of *Z. nevadensis* may be caused by the interspecific transfer of endosymbionts within the gut, but the presence of identical phylotypes in *Trichonympha* species of different termites was quite surprising, because it can be explained only by multiple recent transfers of endosymbionts between flagellates of different termite species. Therefore, we

suspected that the shared phylotypes are artifacts that either resulted from the picking process (*e.g.*, the symbionts stem from other, lysed flagellate cells present in the same gut) or were generated during the pyrosequencing process.

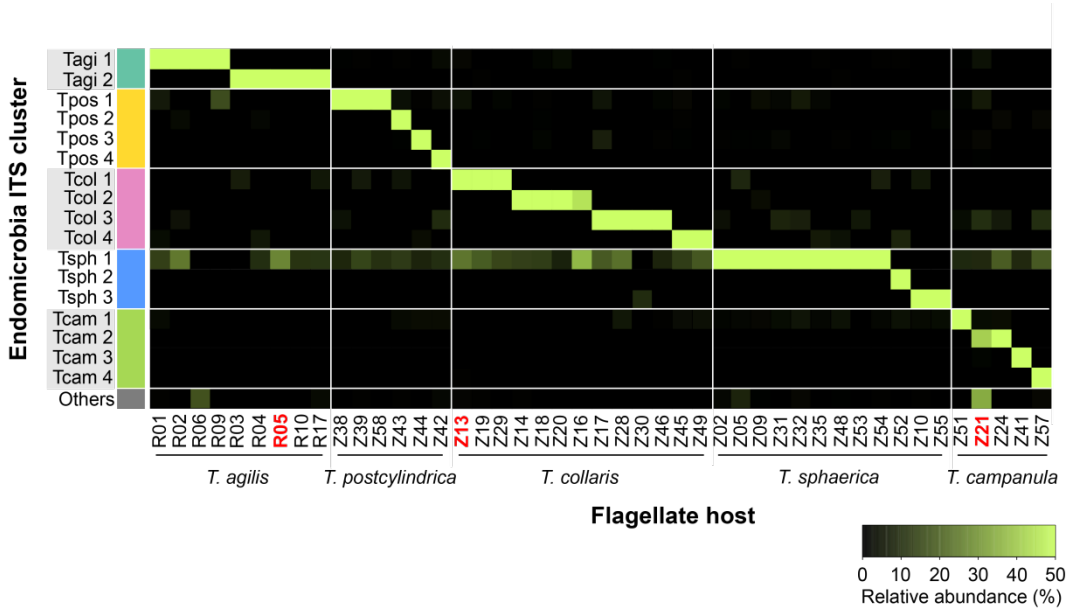

Fig. 2.5. Heatmap of the relative abundance of different phylotypes of Endomicrobia within the pyrotag libraries of all single flagellate cells. The samples were ordered according to the clusters indicated in Fig. 2.2. The samples that were checked for purity by cloning and Sanger sequencing are in red.

To clarify this, we reanalyzed three of the Endomicrobia samples obtained from individual flagellates (R05, Z13 and Z21), which were selected because they had particularly large proportions of suspicious sequences (see Fig. 2.5). To exclude the possibility that the minor phylotypes in the corresponding pyrotag libraries stem from a contamination during the picking process or originated during whole-genome amplification or library preparation (*e.g.*, by contaminated tags), we used aliquots of the samples that had been preserved directly before pyrosequencing (*i.e.*, after the PCR step that introduced the sample-specific tags). The amplicons were ligated into a plasmid vector, and 30 clones of each sample

were sequenced. In each case, all sequences in the library were identical to the major phylotype obtained from the respective flagellate, which clearly identifies the minor phylotypes in the corresponding pyrotag libraries as artifacts generated during the pyrosequencing process.

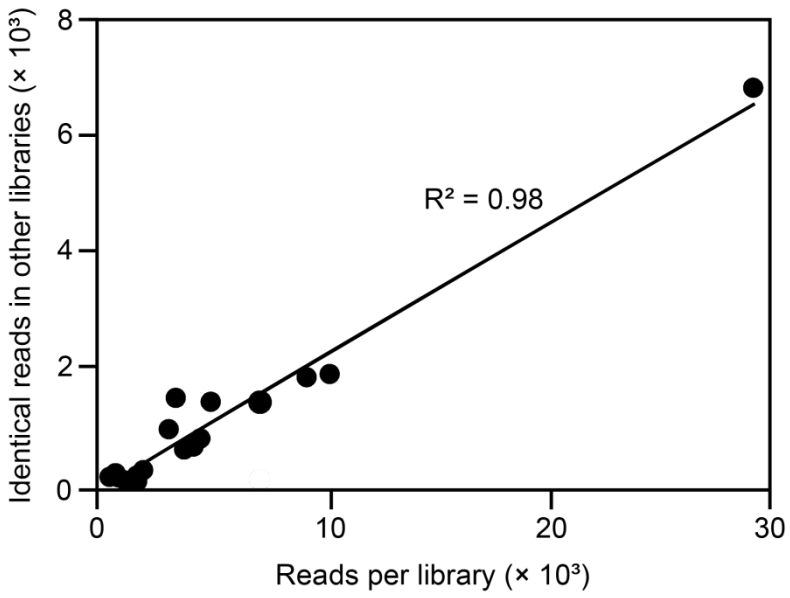

Fig. 2.6. Correlation of the number of reads for the major phylotype in each pyrotag library of individual flagellates to the number of identical reads recovered from the libraries of other flagellate species. Each dot represents the results obtained for an individual sample.

This interpretation is supported by the observation that the reads of Endomicrobia phylotype Tsph1, which was present in most *T. sphaerica* samples and overrepresented in the pyrosequencing run, were recovered in varying abundance from all other flagellate samples (Fig. 2.5; Fig. S2.2). To test this quantitatively, we compared the number of reads for the major phylotype in each flagellate sample to the number of identical reads recovered from samples of other flagellate species (Fig. 2.6). The strong linear correlation corroborates that the minor phylotypes in each sample are not real but artifacts caused by tag switching.

Discussion

Host specificity of Endomicrobia

In this study, we showed that each of the *Trichonympha* cells from both *Z. nevadensis* and *R. santonensis* harbor a single phylotype of "*Candidatus* Endomicrobium trichonymphae". The symbiont populations in each host lineage are phylogenetically distinct even if multiple host species are present in the same gut, corroborating the assumption that the apparent cospeciation between the partners in this symbiosis is caused by vertical (cytoplasmic) inheritance of the endosymbionts (Ikeda-Ohtsubo and Brune 2009).

Although Ikeda-Ohtsubo and Brune (Ikeda-Ohtsubo and Brune 2009) were able to document a congruent topology of the SSU rRNA trees of the symbiotic pairs, the relatively small number of sequences in their clone libraries (10–20 clones) did not allow us to exclude the presence of minor phylotypes of Endomicrobia within each *Trichonympha* cell. Moreover, since the libraries were obtained from flagellate suspensions containing numerous host cells (100–200 cells per sample), it remained possible that individual flagellates harbored a phylotype entirely different from that of the majority of the cells in the suspension.

Prompted by the observation that a large Endomicrobia clone library of *Z. nevadensis* (353 clones) contained a much larger number of 16S rRNA phylotypes than *Trichonympha* species described for this flagellate (Ikeda-Ohtsubo 2007), we reinvestigated the subject. Since the resolution of 16S rRNA analysis was not sufficient due to the high similarity of the sequences, we used the much more variable ITS region, harnessing the advantages of whole-genome amplification and next-generation sequencing technology to achieve a highly resolved analysis of their community structure based on the genomic DNA of the symbiont

populations in individual host cells. Beyond the artifacts inherent to the pyrotag method (see below), it is now safe to conclude that (i) each *Trichonympha* cell harbors a homogeneous population of Endomicrobia symbionts, (ii) *Trichonympha* cells of different species always harbor a phylogenetically distinct symbiont population, and (iii) the phylogenetic diversity of Endomicrobia in the two termites investigated in this study is caused by a microdiversity of *Trichonympha* populations that extends beyond the framework of the currently described species.

It cannot be denied that also the present study reaches its limits when it comes to excluding the possibility of an exchange of symbionts between individuals of the same (or closely related) *Trichonympha* populations present within the same gut. While the *Trichonympha* species in the wood-feeding cockroach *Cryptocercus punctulatus* still reproduce sexually during each molt of the host, this trait has been lost in the *Trichonympha* lineages of termites at an uncertain evolutionary time point (Cleveland 1949). However, artificial feeding of *Z. angusticollis* with the molting hormone 20-hydroxyecdysone still seems to trigger a sexual cycle of their *Trichonympha* flagellates (Messer and Lee 1989), and it is possible that such copulation events still allow a mixing of Endomicrobia phylotypes among host cells of the same species on rare occasions. Nevertheless, the strong and ongoing genome reduction demonstrated in "*Ca.* Endomicrobium trichonymphae" strain Rs-D17 (Hongoh et al. 2008) indicates the absence of genetic exchange and a frequent population bottleneck during the transmission (Wernegreen 2002), a well-documented phenomenon in the intracellular symbionts in the bacteriocytes of aphids (Mira and Moran 2002; Moran et al. 2008).

Endomicrobia populations in *Zootermopsis* and *Reticulitermes*

Assuming that all Endomicrobia species have the same copy number of rRNA genes, the Endomicrobia community in the gut of *Z. nevadensis* is dominated by phylotypes from *T. collaris* (43%) and *T. sphaerica* (31%). These outnumber the diverse populations of "*Ca.* Endomicrobium trichonymphae" associated with *T. postcylindrica* and *T. campanula*, which account for another 19% of the community (Fig. 2.4B). Considering that the colonization density of Endomicrobia differs among the four *Trichonympha* phylotypes in *Z. nevadensis* (Ikeda-Ohtsubo 2007), this is in reasonable agreement with the relative abundance of *Trichonympha* species in this termite reported by Kirby (Kirby 1932), who found that the three species (he did not recognize *T. postcylindrica* as separate from *T. campanula*), were usually present in comparable numbers. The phylogenetic position of the remaining phylotypes (5% of the reads in the library) recovered only from the gut of *Z. nevadensis* but not from picked flagellates (Fig. 2.4A) suggests that they represent Endomicrobia populations of *Trichonympha* lineages that are not included among the picked cells, but whose presence is indicated by the microdiversity of the ITS sequences obtained for each morphotype (Fig. 2.3).

Also in the case of the *T. agilis* cells picked from the gut of *R. santonensis*, the microdiversity of Endomicrobia (Fig. 2.4A) reflects that of the host flagellates (Fig. 2.3), which indicates a strict host specificity of the endosymbionts. Kirby (Kirby 1932) reports that with the exception of *Reticulitermes lucifugus*, all *Reticulitermes* species he investigated harbor only a single species of *Trichonympha*, which is in agreement with the number of SSU rRNA genes of *Trichonympha* species obtained from different *Reticulitermes* species (Fig. 2.1). However, the ITS sequences of picked flagellates reveal the presence of two closely related phylotypes in *R. santonensis* (Fig. 2.3), which is in agreement with

the two phylotypes of "*Ca.* Endomicrobium trichonymphae" obtained from these samples (Fig. 2.4).

The reads obtained for the amplified ITS regions of individual flagellates of *T. campanula* often contained traces of sequence polymorphism, suggesting that the rRNA gene clusters have multiple and slightly divergent copies. So far, there is no genome information for any termite gut flagellate, but the draft genome of the distantly related parabasalid *Trichomonas vaginalis* contains 254 copies of the 18S rRNA gene (Carlton et al. 2007). ITS regions experience low selective constraint and are considered to evolve rapidly by large insertions and/or deletions (Sawada et al. 1997).

Methodological pitfalls of pyrotag sequencing

In amplicon pyrosequencing, the addition of sequence tags to the amplified target sequences allows a parallel processing of numerous samples in the same sequencing run (Hamady et al. 2008). The tags are usually added to both ends of the amplicons because this allows identification of cases of tag switching (Binladen et al. 2007), which result in impossible tags combinations at each terminus and may affect a large proportion of the reads in pyrotag libraries (van Orsouw et al. 2007; Carlsen et al. 2012).

However, tag switching can be recognized and removed during quality control only if both termini of the amplicon are present in the same read. If the length of the amplicon exceeds the read length, a common problem with most sequencing platforms, it is not possible to identify mistagged sequences without any prior information. In this case, they either go entirely unnoticed or are purged from the dataset by omitting all reads that do not pass a certain frequency threshold (*e.g.*, Tai et al. 2015).

Also in this study, the necessities of primer design dictated an amplicon length of about 600 bp (ITS region of Endomicrobia; Fig. 2.2B) that exceeded even the read length of the 454 Titanium technology (up to 400 bp) (Glenn 2011). However, the low diversity of the Endomicrobia community and the highly improbable presence of identical phylotypes in host flagellates of different termites allowed us to identify reads with suspicious tags. Their frequency amounted to 15% of the reads in the entire sequencing run (Fig. S2.2), which is at the upper end of the proportion of mistagged reads reported in previous studies (van Orsouw et al. 2007). The complete absence of these reads from the processed samples prior to the sequencing run excludes that they result from contaminated tags and indicated that they originated during the pyrosequencing step, probably from tag switching during emulsion PCR, as already suggested by other authors (Carlsen et al. 2012). Since the number of mistagged reads in each sample will depend on the number of reads carrying a particular tag (Fig. 2.6; Fig. S2.2), mistagged reads cannot be removed simply by applying a frequency threshold. Particularly in cases that are highly sensitive to tag switching, *e.g.*, the identification of core communities shared across multiple samples, such artifacts have to be excluded using an appropriate experimental design. Otherwise, the simultaneous presence of OTUs in different samples of a pyrotag run has to be interpreted with the necessary caution (Dietrich et al. 2014).

Materials and Methods

Termites

Zootermopsis nevadensis was collected near the Chilao Flats Campground, Angeles National Forest, California, USA in 2006. *Reticulitermes santonensis*, which is synonymous with *Reticulitermes flavipes* (Austin et al. 2005), was

collected near La Gautrelle, Ile d'Oléron, France in 2010. Since then, both species have been maintained in the laboratory on a diet of pine wood and water. Worker termites (pseudergates) were used for all experiments.

Micromanipulation and whole-genome amplification

Termites were dissected and the entire hindgut content was carefully diluted in Solution U (Trager 1934) in Eppendorf tubes. Aliquots of this suspension (10 µL) were transferred to the wells of a Teflon-coated microscope slide and inspected with an inverted microscope with phase-contrast optics (50-fold magnification). Individual *Trichonympha* cells were identified by their morphology and captured using a micropipette attached to a micromanipulator as described before (Thompson et al. 2012), and transferred to a PCR tube with 50 µL Solution U. The tubes were incubated at 95 °C for 10 min to lyse the flagellates, cooled on ice for 2 min, and centrifuged at slow speed (50 × *g*) for 1 min to remove cell debris. The supernatant (including the endosymbionts) was used as template for multiple displacement amplification (MDA) with the REPLI-g UltraFast Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions, except that the reaction time was increased to 4 h.

Library preparation and sequencing

The ITS regions of flagellates and Endomicrobia were amplified by PCR using the MDA products (25-fold diluted) as template and specific primer pairs for the proximal regions of the flanking rRNA genes (Fig. 2.2). Combination of the *Trichonympha*-specific forward primer 18S-Tri-1287f (AAGATTCAC GTAGCTGGG; this study) and the universal reverse primer 28S-1r

(ATGCTTAAATTCAGCGGGT) (Moreira et al. 2007) yielded PCR products of ~700 bp (30 cycles of amplification: 30 s at 94 °C, 30 s at 54 °C, and 60 s at 72 °C), which were directly sequenced as previously described (Thompson et al. 2012).

The Endomicrobia-specific primers 16S-Endo-1502f (AAGGTAGCCGTACGAG A) and 23S-Endo-28r (ACAGTCTTAGCCAAGGCA) were designed on the basis of all Endomicrobia sequences represented in public databases. Both primers were barcoded as described previously (Köhler et al. 2012). Endomicrobia sequences were also amplified directly from undiluted DNA extracted from whole-gut homogenates of *Z. nevadensis* and *R. santonensis* (35 cycles of amplification: 30 s at 94 °C, 30 s at 56 °C, and 60 s at 72 °C). All samples were commercially sequenced in a single sequencing run (454 GS FLX 64 with Titanium technology; GATC Biotech, Konstanz, Germany).

Bioinformatics

The ITS sequences of picked flagellates were aligned *de novo* with *MAFFT* version 7 (Katoh and Standley 2013). After manual curation of the alignment, a neighbor-joining tree was constructed using the *ARB* software suite (Ludwig et al. 2004). The ITS sequences of Endomicrobia were processed as previously described (Dietrich et al. 2014). Briefly, pyrotag reads with a minimum length of 250 bp and a maximum expected error of 0.5 were selected and demultiplexed using their barcode sequences (no mismatch allowed). After removal of barcodes and primer sequences, the sequences were clustered at the 99% similarity level with *UPARSE* (Edgar 2013). Sequences were dereplicated, and representative

phylotypes (most abundant sequence in the respective cluster) were aligned *de novo* using the *MAFFT* aligner in L-INS-I mode with 100 iterations (Katoh and Standley 2013). If necessary, the alignment was manually refined so that all sequences were unambiguously aligned. Maximum-likelihood trees were constructed using *RAxML* version 8.1.3 (Stamatakis 2014) with the 16-state GTR-Γ model and 1,000 bootstraps. The heatmap was generated using the *R* software with the package *heatmap.plus* (Day 2012).

Sequence accession numbers

The ITS region sequences of flagellates have been deposited in GenBank under accession numbers KJ778566–KJ778610. Representative sequences of all Endomicrobia phylotypes obtained from flagellate samples and whole-gut homogenates have been deposited in GenBank under accession numbers KP058245–KP058309.

Acknowledgements

This study was funded by the Max Planck Society. Hao Zheng is a doctoral student in the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC), Marburg. Claire Thompson was funded by a postdoctoral fellowship of the LOEWE Center for Synthetic Microbiology, Marburg. We thank Jared Leadbetter (Caltech) for providing termites.

References

Austin JW, Szalanski AL, Scheffrahn RH, Messenger MT, Dronnet S, Bagnères AG. 2005. Genetic evidence for the synonymy of two *Reticulitermes* species: *Reticulitermes flavipes* and *Reticulitermes santonensis*. *Ann Entomol Soc Am*. 98:395–401.

Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One*. 2:e197.

Brune A, Ohkuma M. 2011. Role of the termite gut microbiota in symbiotic digestion. In: Bignell DE, Roisin Y, Lo N, editors. Biology of Termites: A Modern Synthesis. Dordrecht: Springer, pp. 439–475.

Brune A. 2011. Microbial symbioses in the digestive tract of lower termites. In: Rosenberg E, Gophna U, editors. Beneficial Microorganisms in Multicellular Life Forms. Heidelberg: Springer, pp. 3–25.

Brune A. 2012. Endomicrobia: intracellular symbionts of termite gut flagellates. *J Endocytobiosis Cell Res*. 23:11–15.

Brune A. 2014. Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol*. 12:168–180.

Carlsen T, Aas AB, Lindner D, Vralstad T, Schumacher T, Kauserud H. 2012. Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecol*. 5:747–749.

Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, Wortman JR, Bidwell SL, Alsmark UCM, Besteiro S, et al. 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*. 315:207–212.

Carpenter KJ, Chow L, Keeling PJ. 2009. Morphology, phylogeny, and diversity of *Trichonympha* (Parabasalia: Hypermastigida) of the wood-feeding cockroach *Cryptocercus punctulatus*. *J Eukaryotic Microbiol*. 56:305–313.

Cleveland LR. 1949. Hormone-induced sexual cycles of flagellates. I. gametogenesis, fertilization, and meiosis in *Trichonympha*. *J Morphol*. 85:197–295.

Day A. 2012. *heatmap.plus* – heatmap with more sensible behavior. *R* package version 1.3. http://cran.r-project.org/web/packages/heatmap.plus/index.html.

Dietrich C, Köhler T, Brune A. 2014. The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl Environ Microbiol*. 80:2261–2269.

Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 10:996–998.

García-Martínez, Acinas SG, Antón AI, Rodriguez-Valera F. 1999. Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *J Microbiol Methods*. 36:55–64.

Glenn TC. 2011. Field guide to next-generation DNA sequencers. *Mol Ecol Resour*. 11:759–769.

Grassi B, Sandias A. 1897. The constitution and development of the society of termites: observations on their habits; with appendices on the parasitic protozoa of Termitidae, and on the Embiidae. *Quart J Microsc Sci*. 39:245–332.

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods*. 5:235–237.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Complete genome of the uncultured

Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci U S A*. 105:5555–5560.

Hongoh Y, Ohkuma M. 2010. Termite gut flagellates and their methanogenic and eubacterial symbionts. In: Hackstein JHP, editors. (Endo)Symbiotic Methanogenic Archaea. Heidelberg: Springer, pp. 55–79.

Hongoh Y. 2011. Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell Mol Life Sci*. 68:1311–1325.

Ikeda-Ohtsubo W. 2007. Endomicrobia in termite guts: symbionts within a symbiont. Doctoral thesis, Philipps-Universität Marburg, Marburg, Germany.

Ikeda-Ohtsubo W, Desai M, Stingl U, Brune A. 2007. Phylogenetic diversity of 'Endomicrobia' and their specific affiliation with termite gut flagellates. *Microbiology*. 153:3458–3465.

Ikeda-Ohtsubo W, Brune A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol Ecol*. 18:332–342.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.

Kirby H. 1932. Flagellates of the genus *Trichonympha* in termites. *Univ Calif Publ Zool*. 37:349–476.

Kofoid CA, Swezy O. 1919a. Studies on the parasites of the termites III. On *Trichonympha campanula* sp. nov. *Univ Calif Publ Zool*. 20:41–98.

Kofoid CA, Swezy O. 1919b. Studies on the parasites of the termites IV. On *Leidyopsis sphaerica* gen. nov., sp. nov. *Univ Calif Publ Zool*. 20:99–116.

Köhler T, Dietrich C, Scheffrahn RH, Brune A. 2012. High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher hermites (*Nasutitermes* spp.). *Appl Environ Microbiol*. 78:4691–4701.

Koidzumi M. 1921. Studies on the intestinal protozoa found in the termites of Japan. *Parasitology* 13:235–305.

Leidy J. 1881. The parasites of the termites. *J Acad Nat Sci Philadelphia 2nd Ser*. 8:425–477.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res*. 32:1363–1371.

Messer AC, Lee MJ. 1989. Effect of chemical treatments on methane emission by the hindgut microbiota in the termite *Zootermopsis angusticollis*. *Microb Ecol*. 18:275–284.

Mira A, Moran NA. 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb Ecol*. 44:137–143.

Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet*. 42:165–190.

Moreira D, von der Heyden S, Bass D, López-García P, Chao E, Cavalier-Smith T. 2007. Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Mol Phylogenet Evol*. 44:255–266.

Ohkuma M, Sato T, Noda S, Ui S, Kudo T, Hongoh Y. 2007. The candidate phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol Ecol*. 60:467–476.

Ohkuma M, Noda S, Hongoh Y, Nalepa CA, Inoue T. 2009. Inheritance and diversification of symbiotic trichonymphid flagellates from a common ancestor of termites and the cockroach *Cryptocercus*. *Proc R Soc B*. 276:239–245.

Ohkuma M, Brune A. 2011. Diversity, structure, and evolution of the termite gut microbial community. In: Bignell DE, Roisin Y,Lo N, editors. Biology of Termites: A Modern Synthesis. Dordrecht: Springer, pp. 413–438.

Sato T, Hongoh Y, Noda S, Hattori S, Ui S, Ohkuma M. 2009. *Candidatus* Desulfovibrio trichonymphae, a novel intracellular symbiont of the flagellate *Trichonympha agilis* in termite gut. *Environ Microbiol*. 11:1007–1015.

Sato T, Kuwahara H, Fujita K, Noda S, Kihara K, Yamada A, Ohkuma M, Hongoh Y. 2014. Intranuclear verrucomicrobial symbionts and evidence of lateral gene transfer to the host protist in the termite gut. *ISME J*. 8:1008–1019.

Sawada H, Takeuchi T, Matsuda I. 1997. Comparative analysis of *Pseudomonas syringae* pv. actinidiae and pv. phaseolicola based on phaseolotoxin-resistant ornithine carbamoyltransferase gene (*argK*) and 16S-23S rRNA intergenic spacer sequences. *Appl Environ Microbiol*. 63:282–288.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.

Stingl U, Radek R, Yang H, Brune A. 2005. "*Endomicrobia*": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol*. 71:1473–1479.

Strassert JFH, Köhler T, Wienemann THG, Ikeda-Ohtsubo W, Faivre N, Franckenberg S, Plarre R, Radek R, Brune A. 2012. '*Candidatus* Ancillula trichonymphae', a novel lineage of endosymbiotic *Actinobacteria* in termite gut flagellates of the genus *Trichonympha*. *Environ Microbiol*. 14:3259–3270.

Tai V, James ER, Perlman SJ, Keeling PJ. 2013. Single-cell DNA barcoding using sequences from the small subunit rRNA and internal transcribed spacer region identifies new species of *Trichonympha* and *Trichomitopsis* from the hindgut of the termite *Zootermopsis angusticollis*. *PLoS One*. 8:e58728.

Tai V, James ER, Nalepa CA, Scheffrahn RH, Perlman SJ, Keeling PJ. 2015. The role of host phylogeny varies in shaping microbial diversity in the hindguts of lower termites. *Appl Environ Microbiol*. 81:1059–1070.

Thompson CL, Vier R, Mikaelyan A, Wienemann T, Brune A. 2012. '*Candidatus* Arthromitus' revised: segmented filamentous bacteria in arthropod guts are members of *Lachnospiraceae*. *Environ Microbiol*. 14:1454–1465.

Trager W. 1934. The cultivation of a cellulose-digesting flagellate, *Trichomonas termopsidis*, and of certain other termite protozoa. *Biol Bull*. 66:182–190.

van Orsouw NJ, Hogers RCJ, Janssen A, Yalcin F, Snoeijers S, Verstege E, Schneiders H, van der Poel H, van Oeveren J, Verstegen H, et al. 2007. Complexity reduction of polymorphic sequences (CRoPS^{TM}): A novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One*. 2:e1172.

Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet*. 3:850–861.

Yamin MA. 1979. Flagellates of the orders Trichomonadida Kirby, Oxymonadida Grassé, and Hypermastigida Grassi & Foà reported from lower termites (Isoptera families Mastotermitidae, Kalotermitidae, Hodotermitidae, Termopsidae, Rhinotermitidae, and Serritermitidae) and from the wood-feeding roach Cryptocercus (Dictyoptera: Cryptocercidae). *Sociobiology*. 4:1–119.

Supplementary Information

Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.)

Hao Zheng, Carsten Dietrich, Claire L. Thompson, Katja Meuser, Andreas Brune

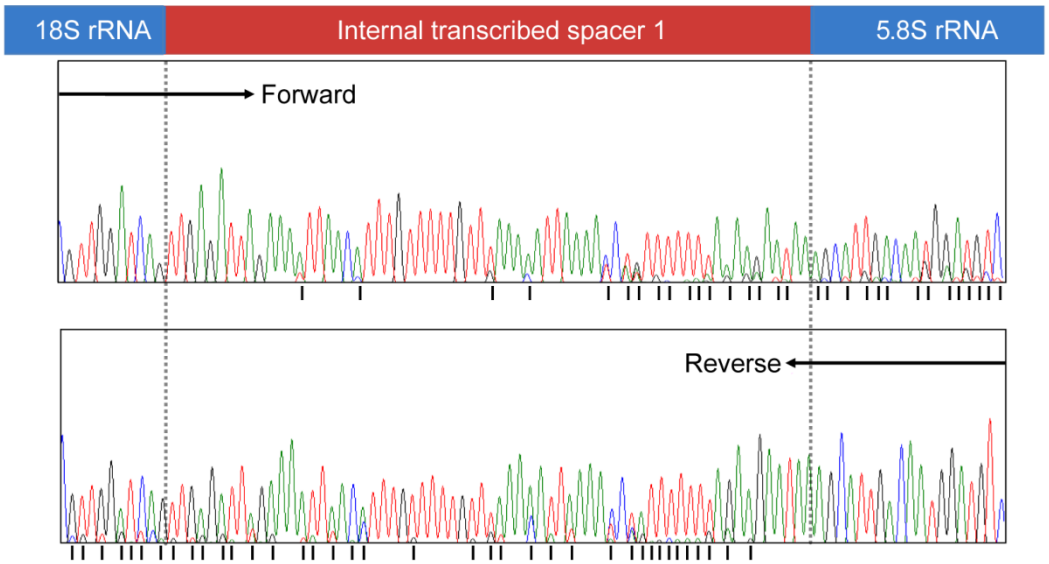

Fig. S2.1. Sequence polymorphism in the ITS region of *Trichonympha campanula*, illustrated by a trace file of the direct Sanger sequencing of the PCR product obtained from a single-flagellate MDA sample (ZnvTrn21). The positions of multiple peaks (indicated by bars) are affected by the sequencing direction but always start in the ITS region, suggesting the presence of both substitutions and indels. For phylogenetic analysis, we used the consensus of the sequences obtained forward and reverse primers, always selecting the stronger signal at the polymorphic positions.

Fig. S2.2. Composition of Endomicrobia ITS libraries obtained by pyrotag analysis of the whole-genome amplification products of individual *Trichonympha* cells. The color code indicates the host species of the major phylotypes. Other phylotypes are shown in grey. The area of each circle represents the number of reads in each library. The names of the samples that were checked for purity by cloning and Sanger sequencing are in red.

Chapter 3

Endomicrobium proavitum, the first isolate of *Endomicrobia* class. nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase

Published online in *Environmental Microbiology* (2015)

doi: 10.1111/1462-2920.12960

Hao Zheng[1], Carsten Dietrich[1], Renate Radek[2], Andreas Brune[1, 3]

[1]Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany; [2]Institute of Biology/Zoology, Free University of Berlin, Berlin, Germany; [3]LOEWE Center for Synthetic Microbiology (SYNMIKRO), Philipps-Universität Marburg, Marburg, Germany

Contributions: H.Z. planned and carried out experiments, analyzed data, discussed results, and wrote the manuscript. C.D. analyzed data, discussed results. R.R. carried out experiments, discussed results. A.B. discussed results, wrote the manuscript, and secured funding.

Abstract

The bacterial tree contains many deep-rooting clades without any cultured representatives. One such clade is “Endomicrobia”, a class-level lineage in the phylum *Elusimicrobia* represented so far only by intracellular symbionts of termite gut flagellates. Here, we report the isolation and characterization of the first free-living member of this clade from sterile-filtered gut homogenate of defaunated (starch-fed) *Reticulitermes santonensis*. Strain Rsa215 is a strictly anaerobic ultramicrobacterium that grows exclusively on glucose, which is fermented to lactate, acetate, hydrogen and CO_2. Ultrastructural analysis revealed a gram-negative cell envelope and a peculiar cell cycle. The genome contains a single set of *nif* genes that encode homologs of Group IV nitrogenases, which were so far considered to have functions other than nitrogen fixation. We documented nitrogenase activity and diazotrophic growth by measuring acetylene reduction activity and $^{15}N_2$ incorporation into cell mass, and demonstrated that transcription of *nifH* and nitrogenase activity occur only in the absence of ammonium. Based on the ancestral relationship to “*Candidatus* Endomicrobium trichonymphae” and other obligate endosymbionts, we propose the name “*Endomicrobium proavitum*” gen. nov., sp. nov. for the first isolate of this lineage and the name “*Endomicrobia*” class. nov. for the entire clade.

Introduction

The intestinal tracts of termites are colonized by a dense microbiota that plays an essential role in symbiotic digestion of lignocellulose and contributes importantly to nitrogen metabolism of the hindgut ecosystem (Brune and Ohkuma, 2011). The bacterial community in the termite hindgut typically comprises hundreds of microbial species, most of which belong to lineages that occur exclusively in this particular habitat (Ohkuma and Brune, 2011). However, many of these lineages have so far resisted cultivation, and most of the knowledge about the microbial ecology of termite guts has been deduced from culture-independent studies (Hongoh, 2011; Brune, 2014a).

A prominent example is the "Endomicrobia" (Brune, 2012), which form a deep-branching clade in the *Elusimicrobia* phylum and make up a substantial portion of the bacterial microbiota in the hindguts of lower termites (Hongoh et al. 2003; Hongoh et al. 2005; Yang *et al.* 2005). Originally referred to as Termite Group 1 (Ohkuma and Kudo, 1996), it was renamed "Endomicrobia" when numerous members of the lineage were recognized as intracellular symbionts of *Trichonympha* spp. and other cellulolytic flagellates (Stingl et al. 2005; Ikeda-Ohtsubo et al. 2007; Ohkuma et al. 2007). The substantial genome corrosion in "*Ca.* Endomicrobium trichonymphae" and simultaneous retention of most pathways for the biosynthesis of amino acids and vitamins (Hongoh et al. 2008) match the obligately intracellular lifestyle of the symbionts and strict cospeciation with their flagellate hosts (Ikeda-Ohtsubo et al. 2007; Ikeda-Ohtsubo and Brune, 2009; Desai et al. 2010; Zheng et al. 2015).

However, members of the Endomicrobia clade have been detected also in the intestinal tract of cockroaches, higher termites, and even ruminants (Ohkuma et al.

2007; Ikeda-Ohtsubo et al. 2010), which all lack the flagellates typical of lower termites. The discovery of novel phylotypes also in defaunated *Reticulitermes santonensis* (Ikeda-Ohtsubo et al. 2010) supported the hypothesis that free-living lineages of Endomicrobia were already part of the gut microbiota before the cockroach ancestor of termites acquired cellulolytic flagellates, whereas intracellular symbiosis evolved at a much later time (Ikeda-Ohtsubo and Brune, 2009).

So far, the only cultivated relative of Endomicrobia is *Elusimicrobium minutum*, a strictly anaerobic ultramicrobacterium that was isolated from filter-sterilized hindgut fluid of a scarab beetle larva and belongs to a different, deep-branching lineage of the *Elusimicrobia* phylum (Geissinger et al. 2009; Brune, 2014b). Inspired by the similarly small diameter of endosymbiotic Endomicrobia (Stingl et al. 2005; Ohkuma et al. 2007) and the observation that the abundance of putatively free-living lineages in lower termites increases upon starch feeding (Ikeda-Ohtsubo et al. 2010), we successfully enriched free-living Endomicrobia strains from filter-sterilized gut homogenates of starch-fed lower termites.

Here, we describe the isolation of *Endomicrobium proavitum* strain Rsa215, the first cultured representative of a new class, *Endomicrobia*. We conducted a detailed physiological and ultrastructural characterization of the strain, focusing on its peculiar cell cycle and its ability to fix dinitrogen with a Group IV nitrogenase. The complete genome of *E. proavitum* has been sequenced (Zheng and Brune 2015); a comparative analysis with the highly reduced genomes of the closely related endosymbionts will be the subject of a separate study.

Results

Isolation

Inoculation of filter-sterilized gut homogenate (0.22-µm pore size) from both starch-fed *R. santonensis* and *Zootermopsis nevadensis* into mineral medium supplemented with glucose yielded enrichment cultures of minute bacteria within 7 days of incubation. Preliminary analysis of the cultures by direct sequencing of the PCR-amplified 16S rRNA genes identified Endomicrobia phylotypes in several cultures, but only enrichment cultures from *R. santonensis* could be successfully transferred. Subsequent deep-agar dilution series in the same medium yielded single colonies of the same phylotype. A pure culture of strain Rsa215 was selected for further characterization.

Phylogenetic position

The 16S rRNA gene sequences of strain Rsa215 and the other phylotypes of Endomicrobia obtained from the enrichment cultures clustered with the clones of putatively free-living Endomicrobia obtained in a previous study (Ikeda-Ohtsubo et al. 2010). The sequence of strain Rsa215 was almost identical to a sequence (clone St414) from a clone library of starch-fed *R. santonensis*. Together with a clone from *Cryptocercus punctulatus*, it formed a sister group of Cluster II of putatively free-living Endomicrobia (Fig. S3.1). The sequence similarities between strain Rsa215 and all other (uncultured) representatives of Endomicrobia, including the endosymbionts of termite gut flagellates and other putatively free-living lineages, ranged from 94% to 96%. The sequences obtained from the enrichment cultures from starch-fed *Z. nevadensis* were phylogenetically distinct and fell into Cluster IV, which also contained another putatively free-living phylotype from starch-fed *R. santonensis* (clone RsStar403; Fig. S3.1). A

comparative analysis of 31 concatenated single-copy genes in the genome of strain Rsa215 (Zheng and Brune 2015) and more than 1,000 other bacterial genomes, including "*Ca*. Endomicrobium trichonymphae" (Hongoh et al. 2008) and *Elusimicrobium minutum* (Herlemann et al. 2009), corroborated the monophyletic nature of the *Elusimicrobia* phylum (Geissinger et al. 2009), and indicated a sister-group position of *Elusimicrobia* and *Spirochaetes* (Fig. 3.1).

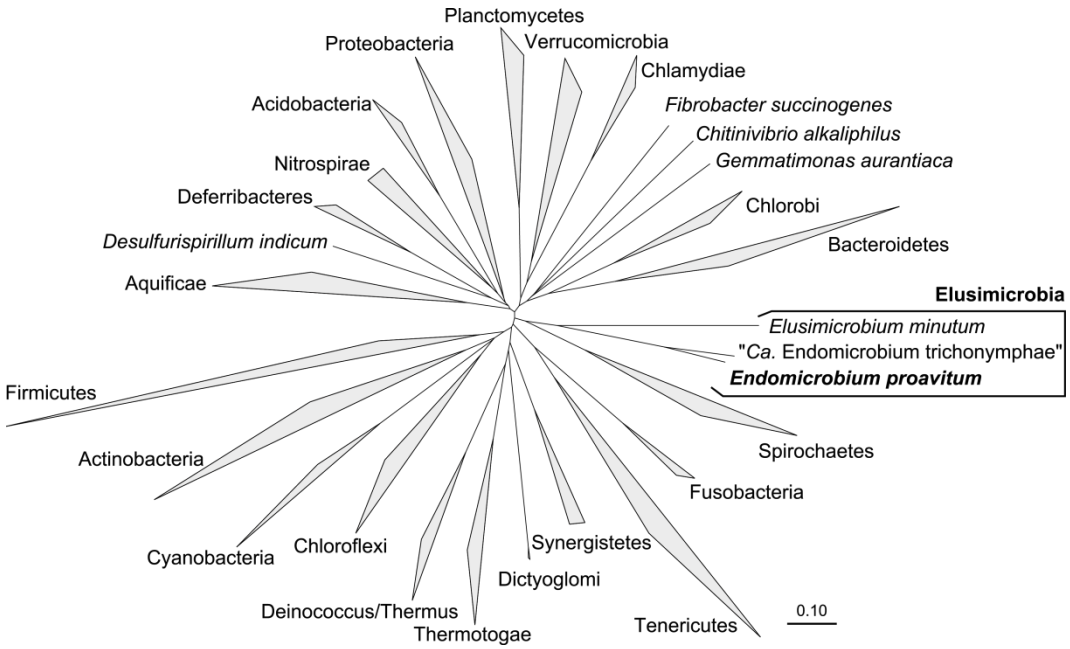

Fig. 3.1. Unrooted maximum-likelihood tree illustrating the phylogenetic position of *Endomicrobium proavitum* strain Rsa215 relative to 1,068 other bacteria with sequenced genomes, based on a concatenated alignment of 31 single-copy genes (6,512 amino acid positions). The monophyly of the *Elusimicrobia* phylum and the sister-group position of *Elusimicrobia* and *Spirochaetes* were highly supported (bootstrap values = 100%). For clarity, the internal nodes of all other phyla were collapsed into wedges.

Cell shape and ultrastructure

Inspection of strain Rsa215 by phase-contrast light microscopy showed both small cocci and thin rods varying in shape and abundance. Electron microscopy confirmed the pleomorphic nature of the cultures. Cells from growing cultures

were mostly rod-shaped or spindle-shaped, with variable diameter (0.15 to 0.30 µm) and length (0.5 to 3.5 µm), often with a distinct swelling at one cell pole (Fig. 3.2A). The diameter of these bud-like structures was two to three times larger than that of the rods and similar to that of the coccoid forms (0.3 to 0.5 µm; Fig. 3.2E and 2F). Moreover, the rod-shaped cells often showed thinner, tubular protrusions reminiscent of hyphae (Fig. 3.2C and 3.2G).

Transmission electron micrographs of ultrathin sections revealed that the cells are surrounded by two membranes. While the inner membrane has the typical appearance of a cytoplasmic membrane, the outer membrane has a trilaminar structure, suggesting the presence of lipopolysaccharide (Nakamura and Mizushima, 1975; Fig. 3.2B). A cell wall is not visible, which is in agreement with the negative Gram staining (not shown). The outer membrane of the rods is corrugated in ultrathin sections (Fig. 3.2D), which matches the granular appearance of the rods in high-resolution scanning electron micrographs (Fig. 3.2A). In the terminal swellings, however, the cell envelope is smoother and resembles that of coccoid forms. The protoplasm contains a central nucleoid (dense filamentous region) and distal electron-translucent regions with electron-dense ribosomes (Fig. 3.2D). We did not observe any flagella or endospores, and coccoid cells lacked cyst-like structures.

The proportion of the different cell forms changed during the growth cycle (Fig. 3.3). Directly after inoculation with an older culture, most of the cells were coccoid (Fig. 3.3A) but turned into rods when growth commenced (Fig. 3.3B). The rods were at their longest during the exponential phase (Fig. 3.3C) and started swelling at one cell pole when the culture approached the stationary phase (Fig. 3.3D). Subsequently, the turbidity of the culture decreased; after 3 weeks, the

culture consisted mostly of coccoid cells and a few short rods without any terminal swelling (Fig. 3.3E).

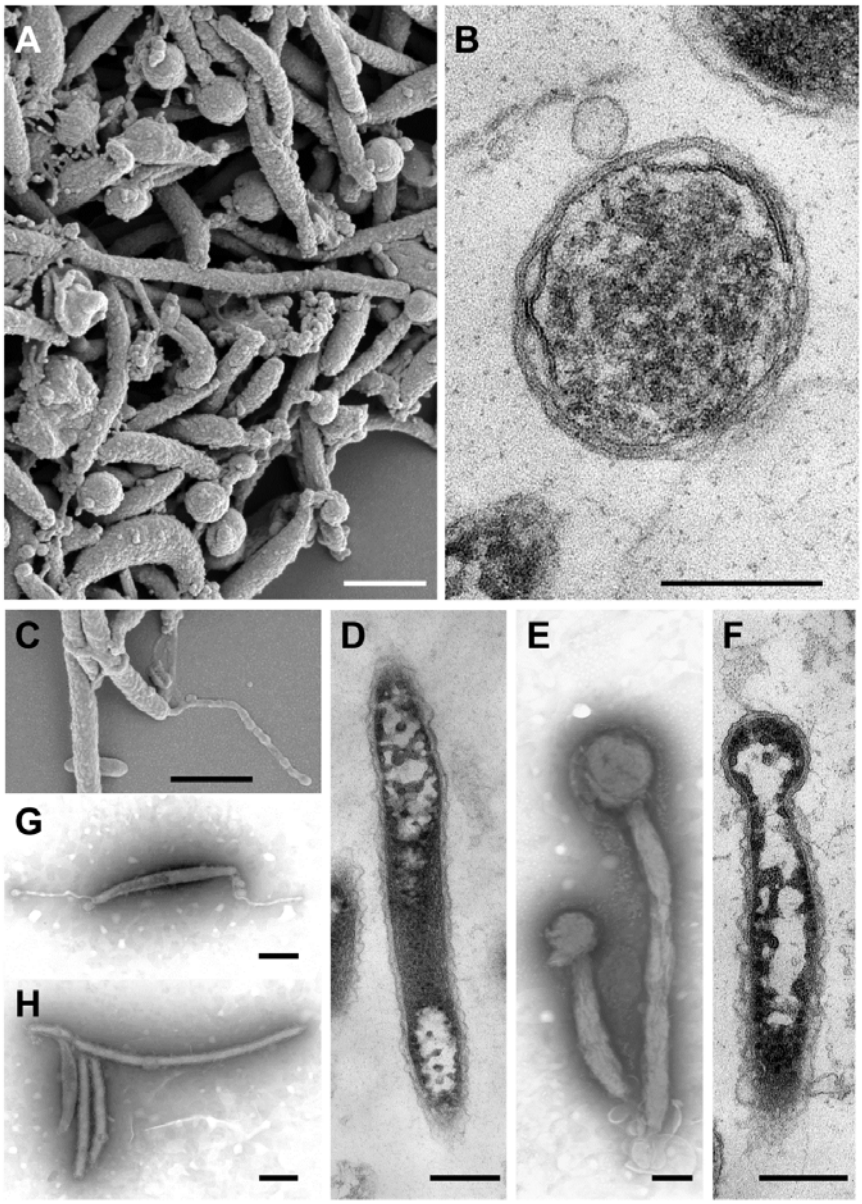

Fig. 3.2. Scanning electron micrographs (A, C) and transmission electron micrographs of ultrathin sections (B, D, F) and negatively stained preparations (E, G, H) of strain Rsa215 in the early stationary phase. Rod-shaped and spindle-shaped cells often showed bud-like swellings at one cell pole (A, E, F) or carried a tubular protrusion (C, G). Thin sections show the presence of an inner and outer membrane (B, D, F). Scale bars are 0.5 µm (A, C, G, H) or 0.2 µm (B, D, E, F).

Physiology of growth

Strain Rsa215 grew exclusively on D-glucose. No growth was observed on glucose derivatives (*N*-acetyl-D-glucosamine, gluconate, 5-ketogluconate, amygdalin, salicin), other monosaccharides (D-fructose, D-galactose, L-rhamnose, L-sorbose, D-tagatose, D-xylose), di- or trisaccharides (D-melibiose, D-saccharose, D-threhalose, D-melezitose, D-raffinose), polysaccharides (glycogen, inulin), sugar alcohols (L-arabitol, D-galactitol, D-ribitol, xylitol), *myo*-inositol, various carboxylic acids (citrate, formate, L-lactate, L-malate, pyruvate), or Casamino acids.

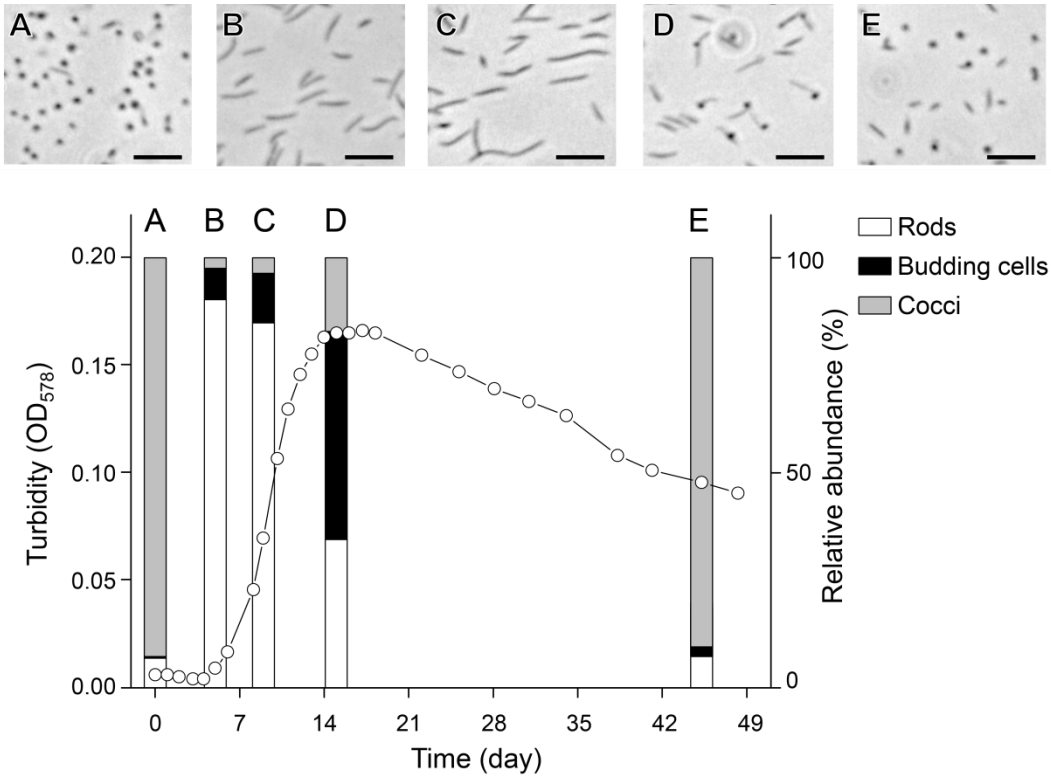

Fig. 3.3. Growth curve of strain Rsa215 in standard medium, and relative abundance of rods, budding cells, and cocci at different growth phases (A–E). The inserts show phase-contrast images of the cultures cells at the respective time points (scale bars, 5 µm).

One glucose was fermented to 0.9 lactate, 0.8 acetate, and 1.8 H_2 (Table 3.1). Carbon and electron recovery were virtually identical and almost complete, which confirms the assumption that acetate and CO_2 were formed in equimolar amounts. When H_2 was added to the headspace at high partial pressure (80 kPa), growth was completely inhibited. The molar growth yield (Table 3.1) was low compared to that of other fermentative bacteria growing on glucose. Doubling time (t_d) in standard medium was 65 h (Fig. 3.3). Growth rate decreased slightly (t_d = 75 h) and growth yield declined by about 30% when yeast extract was omitted, an effect that was even more pronounced during diazotrophic growth (see below). Cultures ceased to grow when transferred to medium lacking vitamin solution. Higher concentrations of yeast extract (0.2%) or the omission of Casamino acids had no effect on growth. Strain Rsa215 grew at pH values between 6.5 and 8.3 (optimum around pH 7) and at temperatures between 16 and 34 °C (optimum at 30 °C); no growth occurred at 37 °C. Strain Rsa215 grew best in freshwater medium but tolerated salt concentrations up to 7 g/l. Strain Rsa215 did not grow in non-reduced medium or in deep-agar tubes incubated under air. The cells were catalase negative.

Table 3.1. Fermentation products, carbon and electron balance, and growth yield of strain Rsa215 in three independent cultures (normalized to one liter).

Glucose consumed (mmol)	Cell mass formed (mg)[a]	Products formed (mmol)			Carbon recovery (%)[b]	Electron recovery (%)[b]	Growth yield (g mol^{-1})
		Lactate	Acetate	H_2			
3.1	33.9	2.8	2.5	5.2			
3.8	36.7	3.5	3.0	6.6	92.3 ± 0.6	93.9 ± 1.4	10.3 ± 0.6
4.0	41.0	3.7	3.2	7.6			

[a]Based on the increase in turbidity and a separately determined conversion factor (in culture tubes, an OD_{578} = 0.1 corresponded to 2.7×10^8 cells ml^{-1} and a dry weight of 26 mg l^{-1}).

[b]Recoveries take into account the assimilated substrate (6.9 mmol glucose per g cell mass, estimated using the formula for *E. coli*; $C_4H_7O_3$); the carbon balance additionally assumes that acetate and CO_2 were formed in equimolar amounts.

Nitrogen fixation

Strain Rsa215 grew on mineral medium (without yeast extract and Casamino acids) in the absence of ammonium only when N_2 was present in the headspace. The growth yield with N_2 was lower than that with ammonium as nitrogen source; no growth occurred when N_2 was replaced with Ar (Fig. 3.4). During growth on N_2, the culture showed acetylene reduction (AR) activity during the exponential phase; no activity was found in the stationary phase (Fig. 3.4).

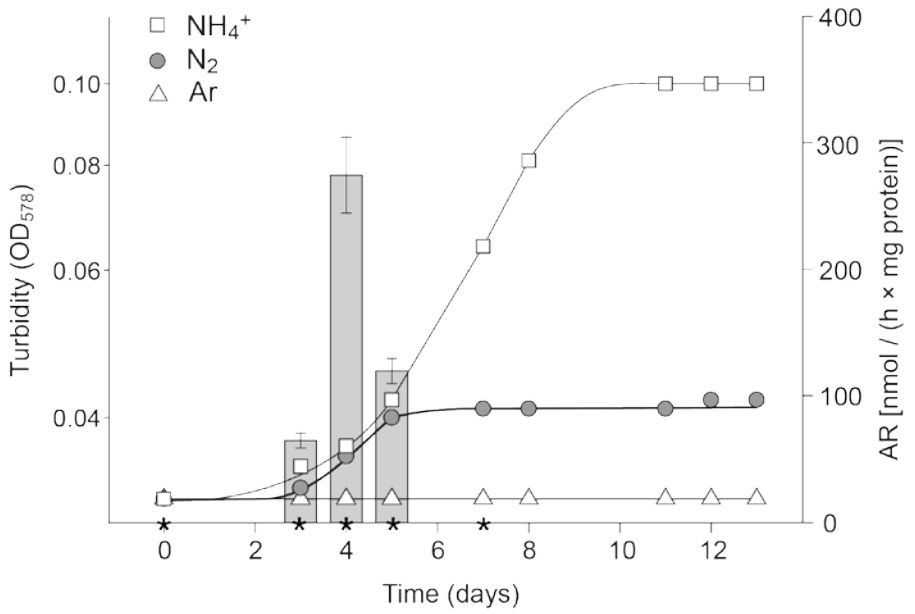

Fig. 3.4. Growth of strain Rsa215 in ammonium-free medium without yeast extracts and Casamino acids, incubated under a headspace of either N_2 or Ar, or supplemented with 5 mM ammonium (NH_4^+), and acetylene reduction (AR) activity (bars) in cultures growing under a headspace of N_2 in the absence of ammonium. Asterisks indicate the time points when AR activity was measured. All values are averages (± SD) of three cultures.

To account for the vitamin solution and any other, unrecognized traces of combined N in the ammonium-free medium, we confirmed the capacity of strain Rsa215 for diazotrophic growth by determining the incorporation of $^{15}N_2$ into cell mass. The proportion of label accumulated by cells grown with $^{15}N_2$ as the sole source of nitrogen indicated that 85% of the nitrogen in cell mass was derived from N_2 (Table 3.2). Assuming that nitrogen fixation occurred exclusively during exponential growth (t_d = 115 h; Fig. 3.4) and that protein and nitrogen constitute 55% and 14% of the cell mass, respectively (dry weight; Neidhart et al. 1990), we estimated a nitrogenase activity of 1.87 µg N_2 h^{-1} (mg protein)$^{-1}$.

Table 3.2. Incorporation of ^{15}N into cell mass by cultures of strain Rsa215 grown in the presence of different nitrogen sources.

Nitrogen source		^{15}N (atom %)		Fixed N (%)[c]
Headspace	Medium	Substrate[a]	Cells[b]	
N_2	–	0.0	0.0 ± 0.0	n.a.[d]
$^{15}N_2$	–	19.6	16.7 ± 0.85	85.2
$^{15}N_2$	NH_4^+	19.6	0.1 ± 0.07	0.5
N_2	$^{15}NH_4^+$	98.0	95.2 ± 1.2	n.a.

[a]N isotopy of the ^{15}N-labeled substrate.

[b]N isotopy of the cell mass at the end of growth. Values are means of two replicate experiments (± mean deviation).

[c]Fraction of N in the cell mass derived from N_2.

[d]Calculation not applicable.

A comparison of this rate with the acetylene reduction rate during exponential growth (Fig. 3.4) resulted in a C_2H_2/N_2 conversion factor of 4.1, which is within the range reported for both purified nitrogenase and cultures of other N_2-fixing bacteria (factor 3–6; Hardy et al. 1973). The presence of ammonium in the medium completely suppressed both acetylene reduction activity (Fig. 3.4) and $^{15}N_2$ fixation (Table 3.2). Transcriptional analysis by RT-PCR with *nifH*-specific primers revealed that the *nifH* gene of strain Rsa215 is transcribed only during diazotrophic growth but not when ammonium was present (Fig. S3.2).

Nitrogenase genes

The complete genome of strain Rsa215 (Zheng and Brune 2015) contains a single gene cluster encoding the structural components of the nitrogenase complex (Fig. 3.5A). Phylogenetic analysis of the genes encoding nitrogenase (*nifD*, *nifK*) and nitrogenase reductase (*nifH*), revealed that the strain possesses a Group IV nitrogenase (Fig. S3.3). In all cases, the genes of strain Rsa215 clustered with homologs from representatives of numerous other bacterial phyla (Cluster IVa; Fig. 3.5B). Closest relatives were homologs from N_2-fixing *Treponema* species isolated from termite guts (Lilburn et al. 2001; Graber et al. 2004), but differences in gene order (Fig. 3.5A) and tree topologies (Fig. S3.3) indicate a rearrangement of individual genes acquired by lateral gene transfer.

Since Group IV nitrogenases are considered Nif-like proteins that do not function in nitrogen fixation (Dos Santos et al. 2012), we analyzed the amino acid sequences in the ligand-binding region of NifD (Fig. 3.5B). Interestingly, the cysteine residues binding the P-cluster ligand and coordinating the Fe atom of the FeMo cofactor of nitrogenases (Groups I–III; Eady, 1996) are conserved also in strain Rsa215 and its relatives in Cluster IVa. Only the histidine at position 442, which coordinates the Mo atom of the FeMo cofactor and is conserved in Group I–III nitrogenases, is replaced with glycine or other residues.

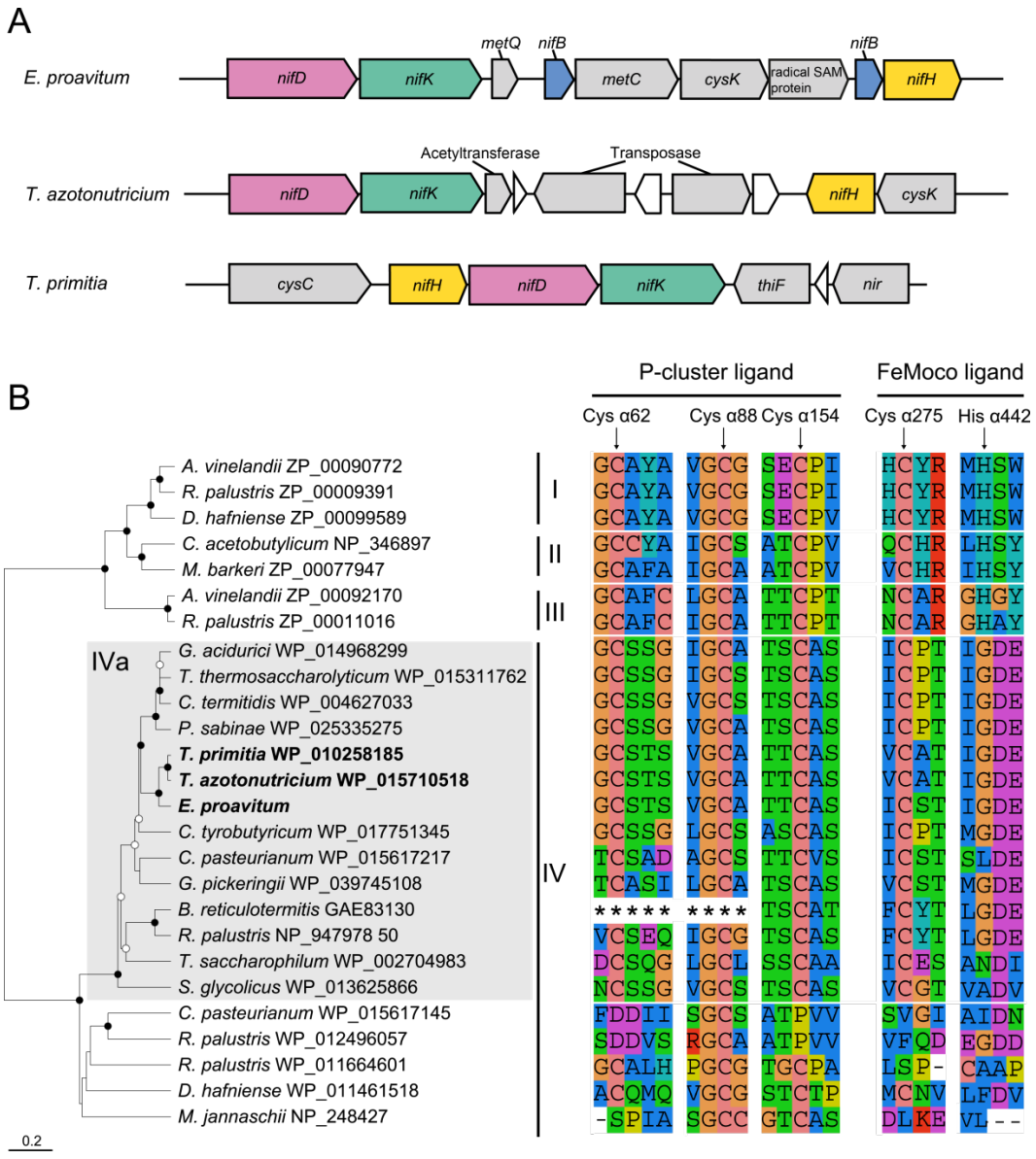

Fig. 3.5. Structure of the gene clusters encoding structural components of Group IV nitrogenases in the genome of *Endomicrobium proavitum* strain Rsa215, *Treponema primitia,* and *Treponema azotonutricium* (A). Phylogenetic tree of the *nifD* genes of strain Rsa215 and the Group IV nitrogenases of other bacterial species, and the conserved amino acids around the crucial residues that coordinate the P-cluster and FeMoco ligands in NifD (B). The full genus names of the taxa are shown in Fig. S3.3. The numbering of the ligands binding P-cluster and FeMoco is based on that of *Azotobacter vinelandii*. The tree is based on deduced amino acid sequences and was rooted using homologs from nitrogenase Groups I–III as outgroup; accession numbers are those of the NCBI protein database). Nodes with high bootstrap values are marked (● > 95%; ○, > 70%; 1,000 replicates).

Discussion

Endomicrobium proavitum strain Rsa215 is the first cultured representative of the candidate class *Endomicrobia*, and the second isolate in the *Elusimicrobia* phylum. Like its distant relative, *Elusimicrobium minutum*, it is a strictly anaerobic ultramicrobacterium with a gram-negative cell wall structure, a distinct cell cycle, and a purely fermentative energy metabolism, which suggests that these are plesiomorphic traits. The ability for diazotrophic growth is based on the presence of a group IV nitrogenase, which documents that at least some of the "*nif*-like" genes from diverse bacterial lineages function in nitrogen fixation. Moreover, its ancestral position to "*Ca.* Endomicrobium trichonymphae" and other uncultured symbionts of termite gut flagellates offers the unique opportunity to study the evolutionary processes that occur during symbiogenesis.

Morphology

Both *E. proavitum* and *Elusimicrobium minutum* qualify as ultramicrobacteria because they can pass through a 0.22-µm membrane filter and have a cell volume < 0.1 μm^3 (Geissinger et al. 2009). Despite the small size of their genomes (*E. proavitum*, 1.59 Mbp; *Elusimicrobium minutum*, 1.64 Mbp) (Herlemann et al. 2009; Zheng and Brune 2015), a substantial fraction of the intracellular volume will be occupied by the nucleoid, which limits the space for ribosomes and other cellular components – arguments that may explain the slow growth and long lag phases of older cultures (discussed in detail already by Geissinger et al. 2009). The reasons for the small size of the two species are unclear; a more efficient

nutrient uptake, as suggested for marine oligotrophs (Dufresne et al. 2003), seems unlikely for species colonizing an intestinal habitat.

Both *E. proavitum* and *Elusimicrobium minutum* (Geissinger et al. 2009) show a distinct cell cycle, with long rods prevailing during the growth phase and coccoid cells or short rods arising during the stationary phase. In the case of *E. proavitum*, the coccoid cells are formed from a bud-like swelling at one of the cell poles. Unlike the microcysts of certain *Proteobacteria*, which have a reinforced cell wall or accumulate refractile intracellular storage materials (e.g., Sutherland and Mackenzie, 1977; Berleman and Bauer, 2004), the coccoid cells differ from the rod-shaped forms only by their enlarged diameter, and it remains to be tested whether they are more resistant to adverse environmental conditions. According to the definition of Hirsch (1974), true budding always serves multiplication, whereas the formation of a resting stage involves the death of the mother cell. The "buds" of *E. proavitum* appear to be formed by unequal cell division and not by hyphal swelling, which occurs in the genus *Hyphomicrobium* (Moore, 1981), and the decrease in turbidity of the cultures during the stationary phase suggests that the coccoid forms increase due to lysis of the rod-shaped cells (Fig. 3.3). Nevertheless, the swelling of the cell pole of the rod-shaped cells may involve the synthesis of a new cell wall, as observed in *Rhodopseudomonas* and *Nitrobacter* spp. (Hirsch, 1974). Moreover, it is not clear whether the tube-like protrusions observed at the poles of the rod-shaped cells are also part of the cellular differentiation process.

Energy metabolism

Endomicrobium proavitum is extremely restricted in its substrate range. Like the distantly related *Elusimicrobium minutum*, it ferments glucose via the Embden-Meyerhoff pathway, but unlike the latter, it uses neither fructose, nor galactose, nor glucosamine as carbon and energy source. Based on the fermentation balance (Table 3.1), taking into account the assimilated substrate and assuming that acetate and CO_2 are formed in equimolar amounts, the stoichiometry of the dissimilatory metabolism can be described by Eqn. 1.

$$\text{Glucose} \rightarrow \text{Lactate} + \text{Acetate} + CO_2 + 2\ H_2 \qquad \text{(Eqn. 1)}$$

The formation of two H_2 per acetate indicates that *E. proavitum* produces one H_2 from NADH (formed during glycolysis) and one H_2 from reduced ferredoxin (formed during pyruvate oxidation), most likely by an electron-confurcating hydrogenase (Schut and Adams, 2009; Huang et al. 2012). The metabolism of *E. proavitum* differs from that of *Elusimicrobium minutum*, which forms substantial amounts of ethanol and alanine but only little lactate and less acetate and H_2 during growth on glucose (Geissinger et al. 2009), and requires both yeast extract and certain amino acids for growth. Since *E. proavitum* grows on mineral medium without the addition of Casamino acids or yeast extract, it must be able to synthesize all amino acids required for protein synthesis, which agrees with the presence of conserved pathways for the biosynthesis of amino acids and cofactors also in the endosymbiotic "*Ca.* Endomicrobium trichonymphae" (Hongoh et al. 2008).

Group-IV nitrogenase

The finding that *E. proavitum* is capable of diazotrophic growth increases the number of bacterial phyla with experimentally verified N_2-fixing representatives to ten (Lilburn et al. 2001; Dos Santos et al. 2012; Wertz et al. 2012; Inoue et al. 2015). It is most interesting that the *nifHDK* genes of *E. proavitum* fall into the phylogenetic radiation of Group IV nitrogenases, which represent a poorly characterized group of "*nif*-like" homologs originally encountered in methanogens and phototrophic bacteria (Raymond et al. 2004). Based on the observation that homologs of *nifD* and *nifH* are widespread in methanogens and expressed constitutively also in the non-nitrogen-fixing *Methanocaldococcus jannaschii*, it has been postulated that they do not function in nitrogen fixation but rather participate in an indispensable and fundamental function(s) in methanogens (Staples et al. 2007). Meanwhile, Group IV nitrogenases have been detected also in numerous other bacterial lineages (Dos Santos et al. 2012), but the simultaneous presence of other nitrogenases never allowed their linkage with diazotrophic growth.

This is the first documentation of a diazotrophic species that fixes nitrogen with a Group IV nitrogenase. The genome of *E. proavitum* contains only a single set of *nifHDK* genes, the nitrogenase of *E. proavitum* is functional, and – as in other diazotrophs (Merrick, 1992) (and in contrast to the *nif*-like homologs of methanogens and phototrophs) – the transcription (of *nifH*) and the expression of nitrogenase genes (AR activity and $^{15}N_2$ incorporation) occur only in the absence of ammonium. The absence of *nifE* and *nifN* genes, whose products function as

scaffolds in the assembly of the FeMo (or FeV) cofactor of nitrogenase (Roll et al. 1995), is not unprecedented among diazotrophs (Mehta and Baross, 2006). The *nifEN* genes probably originated from an ancient duplication of a *nifDK* operon (Fani et al. 2000), and the significant similarity of the products probably allows the cofactor to be assembled directly on the NifD/K tetramer (Howard et al. 2013).

The 4Fe-4S cluster-ligating cysteines and the P-loop/Mg ATP binding motif conserved in the NifH homologs of Group I–III are invariant also in Group IV (including *E. proavitum*; not shown), which is in agreement with the assumption that this protein has a similar function in methanogens and phototrophic bacteria as nitrogenase reductase in diazotrophs (Raymond et al. 2004). The corresponding NifD homologs in methanogens and phototrophic bacteria, however, are highly divergent. Several conserved cysteines involved with P-cluster coordination in Group I–III are missing in Group IV enzymes, and the ligands required to coordinate the FeMo cofactor, the primary cofactor of nitrogenases (Dos Santos et al. 2012), are entirely absent, which indicates a different catalytic role of NifD in these organisms (Staples et al. 2007).

Notably, all cysteine residues required for the coordination of the nitrogenase P-cluster and the Fe atom of the FeMo cofactor are conserved among the NifD homologs in subcluster IVa. The histidine acting as the Mo atom ligand is not conserved, but the nitrogenase activity of *E. proavitum* indicates that this does not preclude production of a functional enzyme. Many members of this subcluster, including isolates from termite gut (Lilburn et al. 2001), are known to be

diazotrophic. In contrast to *E. proavitum*, however, they also possess other nitrogenase gene clusters, and the question whether their Group IVa homologs encode functional nitrogenases remains unresolved.

The demonstration of diazotrophic ability with only group IV nitrogenase in the genome has important implications for our understanding of nitrogen fixation in termite guts. Cultivation-independent studies revealed that the diazotrophic members of the termite gut microbiota are distinct from those of other environments (Yamada et al. 2007). While certain dry-wood termites preferentially express an alternative Mo-independent nitrogenase (*anf*) (Noda et al. 1999; Desai and Brune, 2012), most of the *nifH* homologs from the gut of higher termites have been classified as "pseudo *nifH*" (Ohkuma et al. 1999), later defined as Group IV nitrogenases by Raymond *et al*. (2004). Since some of them were also shown to be expressed (Noda et al. 1999; Desai and Brune, 2012), the contribution of Group IV nitrogenases to the nitrogen economy of the termite has to be reevaluated.

Taxonomy

Strain Rsa215 is a representative of the "putatively free-living Endomicrobia" previously detected in flagellate-free or artificially defaunated termites and cockroaches (Ikeda-Ohtsubo et al. 2010). Its phylogenetically basal position in the Endomicrobia tree (Fig. S3.1) indicates that it shares a common ancestor with the intracellular symbionts. Their high 16S rRNA gene sequence similarity indicates a genus-level relationship between strain Rsa215 and its uncultured relatives from termite guts, but other lineages from insect guts and the bovine rumen are

evidence for additional diversity at the family level. The deep-branching position of the entire clade relative to *Elusimicrobium minutum* (class *Elusimicrobia*) and other uncultured clades (Geissinger et al. 2009) justifies the assignment of the new genus *Endomicrobium* to a new order and class in the *Elusimicrobia* phylum. Interestingly – and in contrast to our previous analysis with a much smaller dataset (Herlemann et al. 2009) – the phylum-level tree indicates that *Elusimicrobia* are a sister group of *Spirochaetes*.

Description of *Endomicrobia* class. nov.

Endomicrobia (En'do.mi.cro'bi.a. N.L. neut. n. *Endomicrobium*, type genus of the type order *Endomicrobiales*; suff. *–ia*, ending to denote a class; N.L. neut. pl. n. *Endomicrobia*, the class of *Endomicrobium*).

The description is the same as for the order *Endomicrobiales* ord. nov.

Type order: *Endomicrobiales* ord. nov.

Description of *Endomicrobiales* ord. nov.

Endomicrobiales (En'do.mi.cro'bi.a'les. N.L. neut. n. *Endomicrobium*, type genus of the order; suff. *–ales*, ending to denote an order; N.L. fem. pl. n. *Endomicrobiales*, the order of *Endomicrobium*). A monophyletic group of bacteria that encompasses *Endomicrobium proavitum* and other, so far uncultivated lineages that occur in the intestinal tract of animals. Type genus: *Endomicrobium* gen. nov.

Description of *Endomicrobiaceae* fam. nov.

Endomicrobiaceae (En'do.mi.cro'bi.a'ce.ae. N.L. neut. n. *Endomicrobium*, type genus of the family; suff. *–aceae*, ending to denote a family; N.L. fem. pl. n.

Endomicrobiaceae, the family of *Endomicrobium*). A monophyletic group of bacteria that encompasses *Endomicrobium proavitum* and other, so far uncultivated lineages that occur in the intestinal tract of insects.

Type genus: *Endomicrobium* gen. nov.

Description of *Endomicrobium* gen. nov.

Endomicrobium (En'do.mi.cro'bi.um. Pref. *endo-* [from Gr. *endon*, within]; N.L. neut. n. *microbium*, microbe; N.L. neut. n. *Endomicrobium*, a genus of microbes that occur within [termite gut flagellates]).

Gram-negative, nonmotile, rod-shaped cells. Obligately anaerobic and catalase negative. Heterotrophic, purely fermentative metabolism. A monophyletic group of bacteria that encompasses *Endomicrobium proavitum* and other, so far uncultivated lineages that occur in the hindgut of termites and cockroaches, either free-living or as intracellular symbionts of gut flagellates.

Type species: *Endomicrobium proavitum* gen. nov., sp. nov.

Description of *Endomicrobium proavitum* sp. nov.

Endomicrobium proavitum (pro.a.vi'tum. L. neut. adj. *proavitum* [from *proavus*, ancestor], belonging to the ancestors; *Endomicrobium proavitum*, a species ancestral to the intracellular symbionts in the genus *Endomicrobium*).

Cultures are pleomorphic and show a distinct cell cycle. Growing cells are rod-shaped or spindle-shaped with variable diameter (0.15 to 0.30 µm) and length (0.5 to 3.5 µm), develop a distinct swelling at one cell pole when approaching the stationary phase. Old cultures consist mostly of coccoid cells (0.3 to 0.5 µm in diameter), and cells are coccoid in the stationary phase. Grows exclusively by

fermentation of D-glucose, with lactate, acetate, CO_2, and hydrogen as end products. No growth on glucose derivatives (*N*-acetyl-D-glucosamine, gluconate, 5-ketogluconate, amygdalin, salicin), other monosaccharides (D-fructose, D-galactose, L-rhamnose, L-sorbose, D-tagatose, D-xylose), di- or trisaccharides (D-melibiose, D-saccharose, D-threhalose, D-melezitose, D-raffinose), polysaccharides (glycogen, inulin), sugar alcohols (L-arabitol, D-galactitol, D-ribitol, xylitol), *myo*-inositol, various carboxylic acids (citrate, formate, L-lactate, L-malate, pyruvate), or Casamino acids. Grows diazotrophically in the absence of ammonium.

Habitat: the hindgut of the termite *Reticulitermes santonensis* (Isoptera: Rhinotermitidae).

Type strain: Rsa215^T (= DSM 29378^T = JCM 30189^T).

Materials and Methods

Enrichment and isolation

Reticulitermes santonensis (synonymous with *Reticulitermes flavipes*; (Austin et al. 2005)) and *Zootermopsis nevadensis* were from the same lab colonies used in a previous study (for details, see Zheng et al. 2015). Worker termites (pseudergates) of both species were placed in Petri dishes and fed with starch powder and water for 5 weeks, as described by Ikeda-Ohtsubo *et al.* (2010). Ten termites each were dissected with fine-tipped forceps, and whole-gut homogenates prepared under anoxic conditions were diluted in 0.5 ml medium (AM-5; see below), as described by Lemke *et al.* (2003). Aliquots of the suspensions were loaded onto centrifuge-tube membrane filters (cellulose acetate, 0.22-µm pore size; Corning® Costar® Spin-X®; Sigma, St. Louis, MO, USA) and centrifuged for 5

min at 5,000 × *g*. The filtrate was used to inoculate enrichment cultures in AM-5 amended with glucose (see below). Pure cultures were isolated from two consecutive deep-agar dilution series in the same medium, as described by Pfennig and Trüper (1981).

Media and cultivation

Cultures were grown in anoxic, bicarbonate-buffered mineral medium (AM-5), which was prepared under a headspace of N_2–CO_2 (80:20, vol/vol), contained a 7-vitamin solution, resazurin (0.4 mg liter^{-1}) as redox indicator, and dithiothreitol (1 mM) as reducing agent, and was adjusted to a final pH of 7.2 (Boga and Brune, 2003). Unless indicated otherwise, the medium was supplemented with yeast extract and Casamino acids (0.1% each). Glucose and other substrates (5 mM) were added from sterile stock solutions; organic acids were added as sodium salts. Growth experiments were routinely conducted in 16-ml rubber-stoppered culture tubes with 5 ml of medium at 30 °C. Growth was determined photometrically following the increase in optical density at 578 nm (OD_{578}). Cell densities were measured with phase-contrast microscopy at 1,600-fold magnification using a Thoma counting chamber. Salt tolerance was determined by mixing the final medium with different volumes of the same medium amended with 4.5% NaCl and 0.5% $MgCl_2$ (mass/volume).

Light and electron microscopy

Bright-field and phase-contrast light microscopy was done with a Zeiss Axiophot microscope (Oberkochen, Germany). For electron microscopy, cultures were fixed in Eppendorf tubes by adding equal volumes of glutaraldehyde solution (5%) in phosphate buffer (0.2 M, pH 7.2) and stored in the fixative for several days. Before each of the following preparation steps, cells were sedimented by centrifugation for 5 min at 4,000 × *g*. For scanning electron

microscopy, cells were washed three times with cacodylate buffer (50 mM, pH 7.2) and postfixed with osmium tetroxide (1%) for 60 min. The cells were washed again three times with distilled water, dehydrated in a graded series of ethanol, dried in hexamethyldisilazane (100%) on cover slips, sputtered with gold, and observed in a SIGMA Field Emission scanning electron microscope (Carl Zeiss, Oberkochen, Germany). For transmission electron microscopy of ultra-thin sections, the fixed cells were washed with phosphate buffer (0.1 M, pH 7.2) three times, postfixed with osmium tetroxide (1%) for 60 min, dehydrated in a graded series of ethanol, and embedded in Spurr's resin (Spurr, 1969). Sections were stained with aqueous, saturated uranyl acetate and lead citrate (Reynolds, 1963) and examined with a Philips EM 208 transmission electron microscope. For negative staining, the fixed cells were pipetted onto a film-coated copper grid and let sediment for 1–2 min. Then, the drop was absorbed with filter paper and substituted by a 1% phosphotungstic acid solution and held for 30–60 s.

Analytical methods

Gram staining followed the procedure of Murray *et al.* (1994); *Escherichia coli* DH5α and *Bacillus subtilis* JH642 were used as controls. Catalase activity was assayed photometrically by monitoring the decomposition of H_2O_2 (Bauer et al. 2000).

Metabolites in the supernatant of fully grown cultures (30 days old) were quantitated by high-performance liquid chromatography using an ion-exclusion column, a UV detector, and a refractive index detector as described previously (Pester and Brune, 2007). Hydrogen formation was analyzed by gas chromatography using a thermal conductivity detector (Pester and Brune, 2007). Electron balances were calculated as previously described (Tholen et al. 2007). Briefly, all metabolites were formally oxidized to CO_2, and the number of

electrons theoretically released from the respective amounts of products was compared with that of the amount of substrate consumed. Expressed on a percent basis, this calculation yielded the electron recovery.

^{15}N incorporation and acetylene reduction assays

Precultures were grown on glucose in ammonium-free AM-5 medium under a headspace of N_2–CO_2 (80:20, vol/vol). Cultures (100 ml) were inoculated with 1 ml preculture in rubber-stoppered glass vials under a headspace (35 ml) of N_2–CO_2. The medium was amended with 7 ml $^{15}N_2$ (98 atom%) in the headspace, or 5 mM $^{15}NH_4Cl$ (98 atom%) from sterile stock solutions (0.5 M). At the end of the growth phase, the cells were harvested by centrifugation, washed with phosphate-buffered saline (NaCl, 137 mM; KCl, 2.7 mM; Na_2HPO_4, 10 mM; KH_2PO_4, 1.8 mM), and the organic N was converted to NH_4^+ using the Kjeldahl method (Lang, 1958). Briefly, 10 mg dry weight of cells were digested in a round-bottom flask (100 ml) with 10 ml 96% sulfuric acid, 5 g K_2SO_4 and 0.1 g $CuSO_4$. The mixture was heated to boiling point until it turned clear (about 1 h). The mixture was cooled to room temperature and neutralized with 10 M NaOH. The N isotopy of NH_4^+ was determined with a GC-IRMS system after conversion to N_2O with sodium hypobromite as described previously (Ngugi and Brune, 2012).

Nitrogenase activity in growing cultures was determined using an acetylene reduction assay. The cells were grown in rubber-stoppered glass vials (100 ml) containing 60 ml liquid culture under a N_2–CO_2 or Ar–CO_2 (80:20, vol/vol) headspace amended with 16% acetylene at 30 °C. Headspace gas (0.1 ml) was sampled and analyzed for ethylene by gas chromatography as described previously (Desai and Brune, 2012).

Phylogenetic analyses

DNA was extracted using the CTAB buffer method (Winnepenninckx et al. 1993). The 16S rRNA genes were amplified using the universal primers 27f and 1492r (Lane, 1991). PCR products were purified and cloned as described previously (Strassert et al. 2010). All clones were commercially sequenced (GATC Biotech, Konstanz, Germany). The 16S rRNA gene sequences were loaded into an existing alignment of Endomicrobia sequences (Ikeda-Ohtsubo et al. 2010) and analyzed using the ARB software suite (version 5.5; Ludwig et al. 2004).

For the genome-based tree, we used a concatenated alignment of 31 single-copy phylogenetic marker genes, which were identified in the genomes using the AMPHORA2 pipeline (Wu and Scott, 2012), which involved Hidden Markov Model searches against the translated ORFs of the respective genomes using the *hmmsearch* binary of *hmmer* (Eddy, 2011) with the recommended *E*-value threshold of 10^{-7}. Sequences were aligned against the seed alignment of the respective Hidden Markov models, filtered with the probabilistic filters supplied with the models (Wu et al. 2012), and trimmed with the reference sequences from the respective marker protein tree. The alignments of the 31 markers were concatenated, and maximum-likelihood trees were inferred by RAxML v8 (Stamatakis, 2014) with the protein-Γ model and the Whelan and Goldman (WAG) substitution matrix.

For the phylogenetic analysis of NifH, NifD and NifK, the corresponding gene sequences obtained by genome analysis (Zheng and Brune 2015) were translated to amino acid sequences and aligned *de novo* with the top BLAST hits and other selected reference sequences from Genbank using the MAFFT aligner (version 7; Katoh and Standley, 2013). After manual refinement of the alignment,

maximum-likelihood trees were constructed using RAxML (version 8; Stamatakis, 2014) with 1,000 bootstraps and the JTT model.

nifH transcription assay

Cultures grown in defined medium in the presence or absence of NH_4Cl (5 mM) were harvested in the exponential phase. Total RNA was extracted using the Ambion® RiboPure™-Bacteria Kit and digested with RNase-free DNase I (Life Technologies, Carlsbad, CA, USA). cDNA was synthesized from 50 ng RNA using SuperScript® III Reverse Transcriptase with Invitrogen® Random Hexamers (Life Technologies). A *nifH* gene fragment (171 bp) was amplified from 30 ng cDNA using PCR with 25 cycles (30 s at 94 °C, 30 s at 56 °C, and 50 s at 72 °C) and the strain-specific primers Epro_nifHf (ATTTTGCGTTGAAGCAGG) and Epro_nifHr (ATAGGCACCGCAAATCCTCC) and separated by gel electrophoresis (2% agarose).

Acknowledgements

This study was funded by the Max Planck Society. Hao Zheng is a doctoral student in the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC), Marburg. We thank Eckhard Völcker, Berlin, for giving us the opportunity to use his SIGMA Field Emission scanning electron microscope.

References

Austin JW, Szalanski AL, Scheffrahn RH, Messenger MT, Dronnet S, Bagnères AG. 2005. Genetic evidence for the synonymy of two *Reticulitermes* species: *Reticulitermes flavipes* and *Reticulitermes santonensis*. *Ann Entomol Soc Am*. 98:395–401.

Bauer S, Tholen A, Overmann J, Brune A. 2000. Characterization of abundance and diversity of lactic acid bacteria in the hindgut of wood- and soil-feeding termites by molecular and culture-dependent techniques. *Arch Microbiol*. 173:126–137.

Berleman JE, Bauer CE. 2004. Characterization of cyst cell formation in the purple photosynthetic bacterium *Rhodospirillum centenum*. *Microbiology*. 150:383–390.

Boga HI, Brune A. 2003. Hydrogen-dependent oxygen reduction by homoacetogenic bacteria isolated from termite guts. *Appl Environ Microbiol*. 69:779–786.

Brune A, Ohkuma M. 2011. Role of the termite gut microbiota in symbiotic digestion. In: Bignell DE, Roisin Y ,Lo N, editors. Biology of Termites: A Modern Synthesis. Dordrecht: Springer, pp. 439–475.

Brune A. 2012. Endomicrobia: intracellular symbionts of termite gut flagellates. *J Endocytobiosis Cell Res*. 23:11–15.

Brune A. 2014a. Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol*. 12:168–180.

Brune A. 2014b. The family *Elusimicrobiaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E ,Thompson F, editors. The Prokaryotes. Berlin: Springer, pp. 637–640.

Desai MS, Strassert JFH, Meuser K, Hertel H, Ikeda-Ohtsubo W, Radek R, Brune A. 2010. Strict cospeciation of devescovinid flagellates and *Bacteroidales* ectosymbionts in the gut of dry-wood termites (Kalotermitidae). *Environ Microbiol*. 12:2120–2132.

Desai MS, Brune A. 2012. *Bacteroidales* ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. *ISME J*. 6:1302–1313.

Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R. 2012. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics*. 13:162.

Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V, Duprat S, Galperin MY, Koonin EV, Le Gall F, et al. 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci U S A*. 100:10020–10025.

Eady RR. 1996. Structure-function relationships of alternative nitrogenases. *Chem Rev*. 96:3013–3030.

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol*. 7:e1002195.

Fani R, Gallo R, Lio P. 2000. Molecular evolution of nitrogen fixation: the evolutionary history of the *nifD*, *nifK*, *nifE*, and *nifN* genes. *J Mol Evol*. 51:1–11.

Geissinger O, Herlemann DPR, Möschel E, Maier UG, Brune A. 2009. The ultramicrobacterium "*Elusimicrobium minutum*" gen. nov., sp. nov., the first cultivated representative of the Termite Group 1 phylum. *Appl Environ Microbiol*. 75:2831–2840.

Graber JR, Leadbetter JR, Breznak JA. 2004. Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl Environ Microbiol*. 70:1315–1320.

Hardy RWF, Burns RC, Holsten RD. 1973. Applications of the acetylene-ethylene assay for measurement of nitrogen fixation. *Soil Biol Biochem*. 5:47–81.

Hartmann LS, Barnum SR. 2010. Inferring the evolutionary history of Mo-dependent nitrogen fixation from phylogenetic studies of *nifK* and *nifDK*. *J Mol Evol*. 71:70–85.

Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P, Brune A. 2009. Genomic analysis of "*Elusimicrobium minutum*," the first cultivated representative of the phylum "*Elusimicrobia*" (formerly Termite Group 1). *Appl Environ Microbiol*. 75:2841–2849.

Hirsch P. 1974. Budding bacteria. *Annu Rev Microbiol*. 28:391–444.

Hongoh Y, Ohkuma M, Kudo T. 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). *FEMS Microbiol Ecol*. 44:231–242.

Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N, Kudol T. 2005. Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Appl Environ Microbiol*. 71:6590–6599.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci U S A*. 105:5555–5560.

Hongoh Y. 2011. Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell Mol Life Sci*. 68:1311–1325.

Howard JB, Kechris KJ, Rees DC, Glazer AN. 2013. Multiple amino acid sequence alignment nitrogenase component 1: insights into phylogenetics and structure-function relationships. *PLoS One*. 8:e72751.

Huang H, Wang S, Moll J, Thauer RK. 2012. Electron bifurcation involved in the energy metabolism of the acetogenic bacterium *Moorella thermoacetica* growing on glucose or H_2 plus CO_2. *J Bacteriol*. 194:3689–3699.

Ikeda-Ohtsubo W, Desai M, Stingl U, Brune A. 2007. Phylogenetic diversity of 'Endomicrobia' and their specific affiliation with termite gut flagellates. *Microbiology*. 153:3458–3465.

Ikeda-Ohtsubo W, Brune A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol Ecol*. 18:332–342.

Ikeda-Ohtsubo W, Faivre N, Brune A. 2010. Putatively free-living '*Endomicrobia*' – ancestors of the intracellular symbionts of termite gut flagellates? *Environ Microbiol Rep*. 2:554–559.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.

Lane DJ. 1991. 16S/23S rRNA sequencing. In: Stackebrandt E ,Goodfellow M, editors. Nucleic Acids Techniques in Bacterial Systematics. Chichester: John Wiley & Sons, pp. 115–175.

Lang CA. 1958. Simple microdetermination of Kjeldahl nitrogen in biological materials. *Anal Chem*. 30:1692–1694.

Lemke T, Stingl U, Egert M, Friedrich MW, Brune A. 2003. Physicochemical conditions and microbial activities in the highly alkaline gut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Appl Environ Microbiol*. 69:6650–6658.

Lilburn TG, Kim KS, Ostrom NE, Byzek KR, Leadbetter JR, Breznak JA. 2001. Nitrogen fixation by symbiotic and free-living spirochetes. *Science*. 292:2495–2498.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res*. 32:1363–1371.

Mehta MP, Baross JA. 2006. Nitrogen fixation at 92°C by a hydrothermal vent archaeon. *Science*. 314:1783–1786.

Merrick MJ. 1992. Regulation of nitrogen fixation genes in bacteria. In: Stacey G, Burris RH ,Evans HJ, editors. Biological Nitrogen Fixation. New York, USA: Chapman and Hall, pp. 835–876.

Moore RL. 1981. The biology of *Hyphomicrobium* and other prosthecate, budding bacteria. *Annu Rev Microbiol*. 35:567–594.

Murray RGE, Doetsch RN, Fobinow CF. 1994. Determinative and cytological light microscopy. In: Gerhardt P, Murray RGE, Wood WA ,Krieg NR, editors. Methods for General and Molecular Bacteriology. Washington, D.C.: American Society for Microbiology, pp. 21–41.

Nakamura K, Mizushima S. 1975. In vitro reassembly of the membranous vesicle from Escherichia coli outer membrane components. Role of individual components and magnesium ions in reassembly. *Biochim Biophys Acta*. 413:371–393.

Neidhart FC, Ingraham JI, Schaechter M. 1990. Physiology of the Bacterial Cell: A Molecular Approach. Sunderland, USA: Sinauer Associates.

Ngugi DK, Brune A. 2012. Nitrate reduction, nitrous oxide formation, and anaerobic ammonia oxidation to nitrite in the gut of soil-feeding termites (*Cubitermes* and *Ophiotermes* spp.). *Environ Microbiol*. 14:860–871.

Noda S, Ohkuma M, Usami R, Horikoshi K, Kudo T. 1999. Culture-independent characterization of a gene responsible for nitrogen fixation in the symbiotic microbial community in the gut of the termite *Neotermes koshunensis*. *Appl Environ Microbiol*. 65:4935–4942.

Ohkuma M, Kudo T. 1996. Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. *Appl Environ Microbiol*. 62:461–468.

Ohkuma M, Noda S, Kudo T. 1999. Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Appl Environ Microbiol*. 65:4926–4934.

Ohkuma M, Sato T, Noda S, Ui S, Kudo T, Hongoh Y. 2007. The candidate phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol Ecol*. 60:467–476.

Ohkuma M, Brune A. 2011. Diversity, structure, and evolution of the termite gut microbial community. In: Bignell DE, Roisin Y ,Lo N, editors. Biology of Termites: A Modern Synthesis. Dordrecht: Springer, pp. 413–438.

Pester M, Brune A. 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J*. 1:551–565.

Pfennig N, Trüper HG. 1981. Isolation of members of the families *Chromatiaceae* and *Chlorobiaceae*. In: Starr MP, Stolp H, Trüper HG, Balows A ,Schlegel HG, editors. The Prokaryotes. Berlin: Springer-Verlag, pp. 229–289.

Raymond J, Siefert JL, Staples CR, Blankenship RE. 2004. The natural history of nitrogen fixation. *Mol Biol Evol*. 21:541–554.

Reynolds ES. 1963. The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *J Cell Biol*. 17:208–212.

Roll JT, Shah VK, Dean DR, Roberts GP. 1995. Characteristics of NIFNE in *Azotobacter vinelandii* strains. *J Biol Chem*. 270:4432–4437.

Schut GJ, Adams MW. 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol*. 191:4451–4457.

Spurr AR. 1969. A low-viscosity epoxy resin embedding medium for electron microscopy. *J Ultrastruct Res*. 26:31–43.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.

Staples CR, Lahiri S, Raymond J, Von Herbulis L, Mukhophadhyay B, Blankenship RE. 2007. Expression and association of group IV nitrogenase

NifD and NifH homologs in the non-nitrogen-fixing archaeon *Methanocaldococcus jannaschii*. *J Bacteriol*. 189:7392–7398.

Stingl U, Radek R, Yang H, Brune A. 2005. "*Endomicrobia*": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol*. 71:1473–1479.

Strassert JFH, Desai MS, Radek R, Brune A. 2010. Identification and localization of the multiple bacterial symbionts of the termite gut flagellate Joenia annectens. *Microbiology*. 156:2068–2079.

Sutherland IW, Mackenzie CL. 1977. Glucan common to the microcyst walls of cyst-forming bacteria. *J Bacteriol*. 129:599–605.

Tholen A, Pester M, Brune A. 2007. Simultaneous methanogenesis and oxygen reduction by *Methanobrevibacter cuticularis* at low oxygen fluxes. *FEMS Microbiol Ecol*. 62:303–312.

Winnepenninckx B, Backeljau T, De Wachter R. 1993. Extraction of high molecular weight DNA from molluscs. *Trends Genet*. 9:407.

Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics*. 28:1033–1034.

Wu M, Chatterji S, Eisen JA. 2012. Accounting for alignment uncertainty in phylogenomics. *PLoS One*. 7:e30288.

Yamada A, Inoue T, Noda S, Hongoh Y, Ohkuma M. 2007. Evolutionary trend of phylogenetic diversity of nitrogen fixation genes in the gut community of wood-feeding termites. *Mol Ecol*. 16:3768–3777.

Yang H, Schmitt-Wagner D, Stingl U, Brune A. 2005. Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ Microbiol*. 7:916–932.

Zheng H, Brune A. 2015. Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum *Elusimicrobia*). *Genome Announc*. 3:e00679-15.

Zheng H, Dietrich C, Thompson CL, Meuser K, Brune A. 2015. Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microbes Environ*. 30:92–98.

Supplementary Information

Endomicrobium proavitum, the first isolate of _Endomicrobia_ class. nov. (phylum _Elusimicrobia_) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase

Hao Zheng, Carsten Dietrich, Renate Radek, Andreas Brune

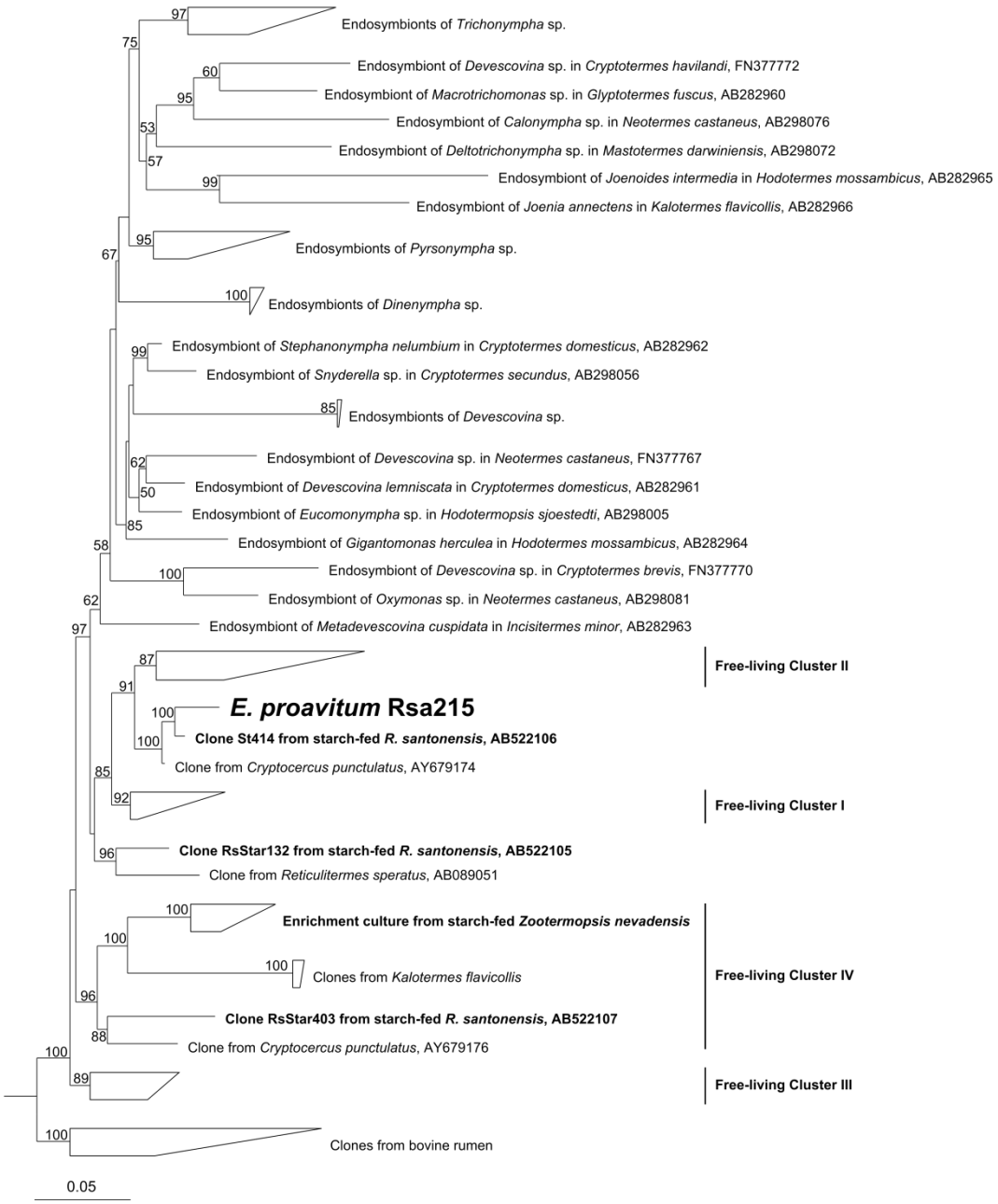

Fig. S3.1. Phylogenetic tree of *Endomicrobia*, showing the phylogenetic position of Endomicrobium proavitum strain Rsa215 and 16S rRNA sequences obtained from starch-fed termites (clone libraries and enrichment cultures; shown in bold) relative to the flagellate endosymbionts and several clusters of putatively free-living phylotypes previously identified by Ikeda-Ohtsubo et al. (2010). The tree was reconstructed by maximum-likelihood analysis of an alignment of 16S rRNA sequences sequences (1064 positions) and rooted with *Elusimicrobium minutum* as outgroup. Numbers at nodes indicate bootstrap support support for 1,000 replicates (only values > 50 are shown).

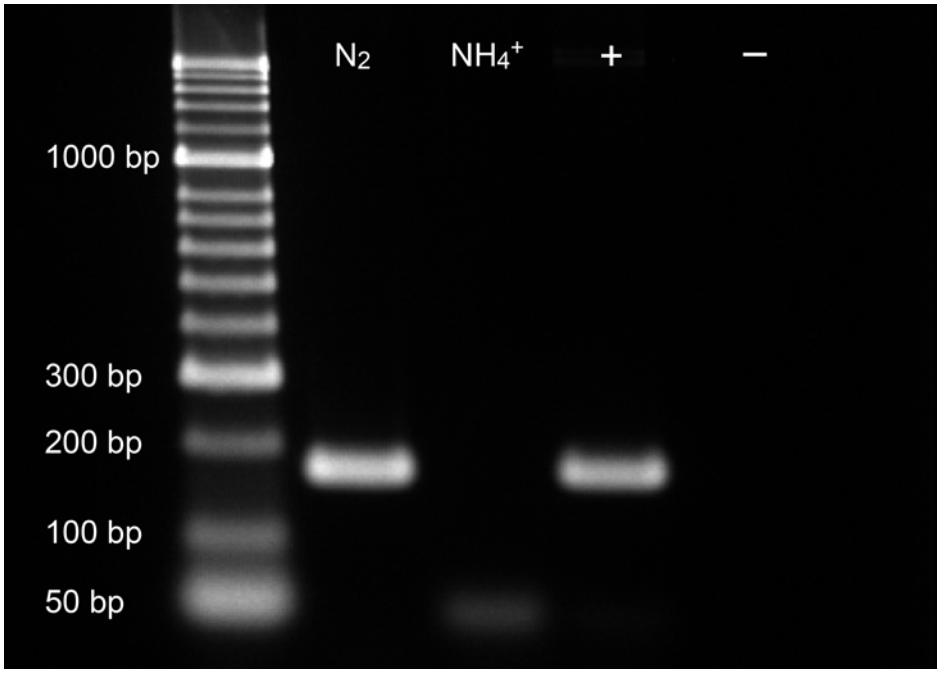

Fig. S3.2. Electropherogram of RT-PCR products of *nifH* transcripts obtained from cultures of strain Rsa215 growing under N_2-fixing condition (N_2) or in the presence of ammonium (NH_4^+). cDNA was amplified templates from total RNA with strain-specific *nifH* primers. DNA of strain Rsa215 served as positive (+) control, and reactions without reverse transcriptase served as negative (–) control.

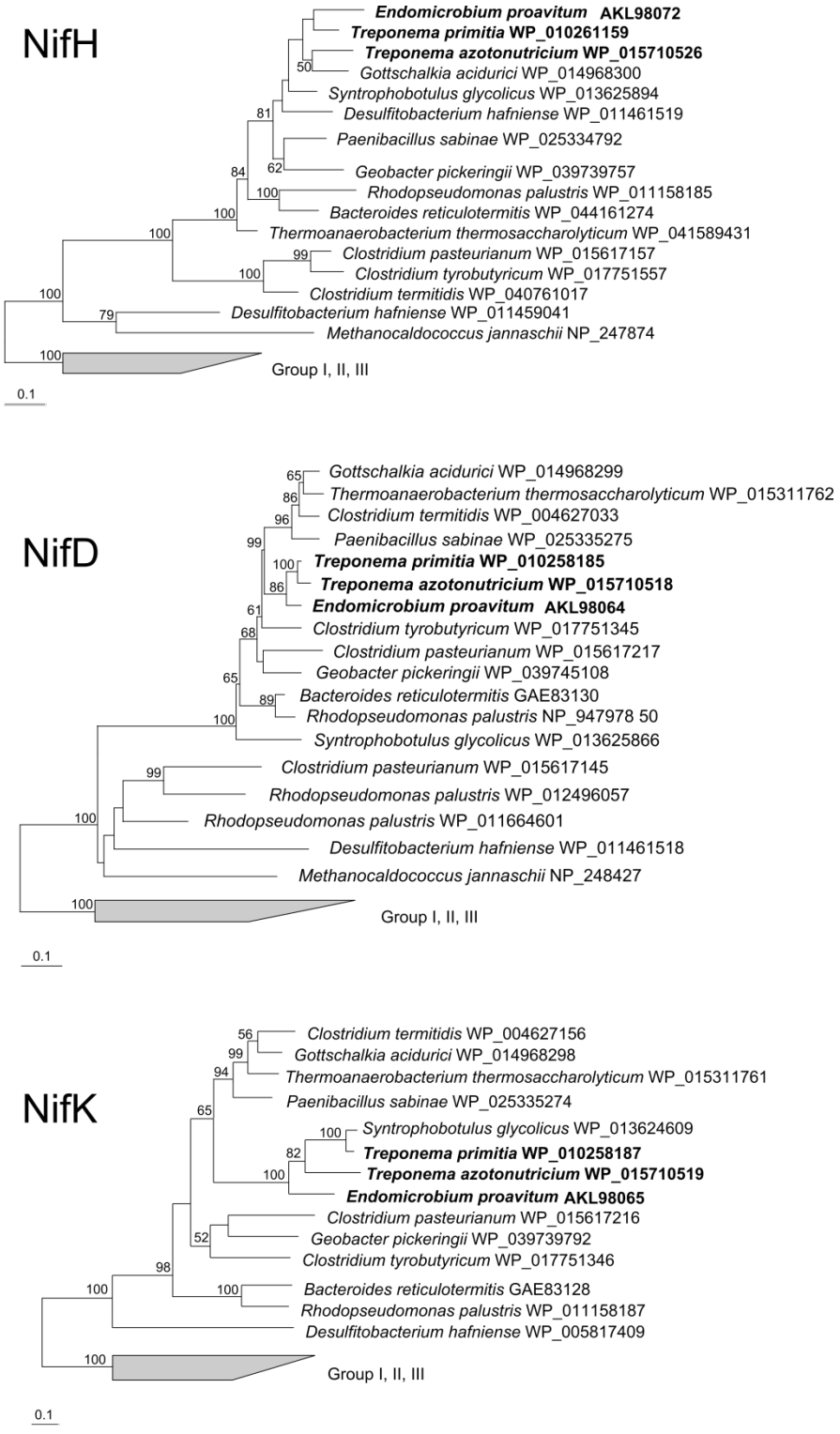

Fig. S3.3. Maximum-likelihood trees of the structural genes of Group IV nitrogenases of *Endomicrobium proavitum* strain Rsa215 and other bacterial species. The trees are based on deduced amino acid sequences and were rooted using homologs from nitrogenase Group I–III as outgroup. Accession numbers are for the NCBI protein database. Nodes with bootstrap values > 50% are marked (1,000 replicates).

Chapter 4

Loss and gain in *Endomicrobia* genomes: From free living bacteria to intracellular symbionts

Manuscript in preparation

Hao Zheng, Carsten Dietrich, Andreas Brune

Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

Contributions: H.Z. planned and carried out experiments, analyzed data, discussed results, and wrote the manuscript. C.D. analyzed data, discussed results. A.B. discussed results and secured funding.

Abstract

Intracellular mutualistic symbiosis is an important strategy in ecology. During the transition of lifestyle, the genomes of endosymbionts undergo degenerations, which also allow us to investigate the evolution of the bacterial genomes. This study presents the comparative analysis of the genome of termite gut flagellate endosymbionts '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17 and that of the gut free-living relative *Endomicrobium proavitum*. The results show that both genomes share many homologs, while all disrupted pathways remain intact in *E. proavitum*. Besides the deletions of genes, the endosymbiont seems to obtain heterologous genes to alter its metabolism strategy. The genome of endosymbionts shows a function-independent purifying selection among all the protein-coding genes. The endosymbiosis between *Endomicrobia* and termite gut flagellate provides a particular model for the study of genome evolution in relative recent endosymbionts.

Introduction

Symbiosis is an important mechanism in ecology and evolution for many organisms belonging to different taxa (de Bary 1879). Many eukaryotic cells possess intracellular symbionts in their cytoplasm, which is sometimes essential for the viability. Intracellular mutualistic bacteria are notable among cellular life forms because they maintain extremely small genomes (Moran and Bennett 2014). Numerous studies have focused on the evolution mechanism of genome reduction during the endosymbiosis progress (McCutcheon and Moran 2012).

Wood-feeding lower termites harbor symbiotic gut flagellates that support the termite nutritionally (Brune 2011) and these flagellates themselves host specific intracellular symbiotic bacteria (Hongoh and Ohkuma 2010). The relationship between termites and cellulolytic flagellates in their hindgut is a well-known system of symbiosis (Brune and Dietrich 2015). These triplex symbiotic relationships involving the three domains of life are potential excellent models for the study of symbioses between bacteria and eukaryotes.

Endomicrobia are endosymbiotic bacteria that represent a deep-branching lineage in the phylum *Elusimicrobia*, which are only found exclusively abundant in the termite gut environment. Cospeciation analysis revealed a strictly vertical transmission of *Endomicrobium* symbionts by their *Trichonympha* hosts (Zheng et al. 2015a). The complete genome of the endosymbionts '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17 revealed that it has a relatively smaller genome and many pathways have been interrupted by pseudogenes (Hongoh et al. 2008). However, there was no genomic information from free-living ancestral relative, which blocks the detailed analysis of the mechanism of genome evolution in *Endomicrobia*. Recently, we obtained the first isolate of the

Endomicrobia clade, *Endomicrobium proavitum*, which is a gut free-living strain from the same termite genus (*Reticulitermes*) as the endosymbionts (Zheng et al. 2015b). The complete genome of *E. proavitum* was also sequenced (Zheng and Brune 2015). Phylogenies based on the 16S rRNA gene and ribosomal protein-coding genes analyses consistently show that the free-living isolate is closely related to the intracellular symbiont. Thus, the comparative analysis of the genomes from both endosymbionts and its ancestral free-living would allow us to study the mechanism of the genome reduction in this symbiosis system.

In this study, we provide the complete genomic sequence of *E. proavitum* isolated from the termite gut and present a comparative analysis of the two completed genomes of both endosymbionts and free-living strain, addressing the possible mechanism of the genome degeneration during the endosymbiosis association.

Results

Genome features

E. proavitum has a relatively small circular chromosome (1.59 Mbp; Table 1), with an average GC content of 39.3 mol%. The genome contains 1,406 predicted genes, of which 1,341 (95.7%) code for proteins designated to different cluster of orthologous group of proteins (COGs; Fig. 4.1), 46 (3.1%) code for RNA genes, and 19 (1.3%) are pseudogenes. The genome contains only a single rRNA operon, which is the same as other two genomes in the phylum *Elusimicrobia* and in agreement with the long doubling time of the organism (Zheng et al. 2015b). The detailed annotation of all protein-coding genes and their COG assignments are presented in the supplemental material (Table S4.1).

Table 4.1. Features of the genomes

	Size (Mbp)	G+C content (%)	Number of CDS[a]	Average length of CDS (bp)	tRNA genes	rRNA operons	Pseudo-genes
E. proavitum	1.59	39.3	1341	1066	46	1	19
Ca. E. trichonymphae	1.13	35.2	761	984	45	1	121
E. minutum	1.64	39.9	1529	951	45	1	20

[a]Excluding pseudogenes.

Metabolic reconstruction of the *E. proavitum* genome

Glucose-6-phosphate (G6P) transporter. Pure culture of *E. proavitum* can only utilize glucose as carbon and energy source, and produces H_2, CO_2, lactate, and acetate as major fermentation products (Zheng et al. 2015b). Unlike the free-living strain, which imports glucose via phosphotransferase system (PTS) present in the genome (Fig. 4.2), the endosymbiotic strain Rs-D17 use UhpC transporter to obtain G6P as the main energy source. The *uhpC* gene is absent in *E. proavitum* genome, and the phylogenetic analysis of UhpC amino acid sequences shows that the sequence of the endosymbiotic strain are closely related to that of *E. minutum* and those from *Clostridium* (Fig. S4.1) indicating this gene is either acquired latterly by horizontally gene transfer (HGT) or lost by the free-living relative during the process of evolution. The resulting sugar phosphates are converted to fructose 6-phosphate and degraded to pyruvate via the Embden-Meyerhof pathway. Pyruvate is converted to lactate by lactate dehydrogenase and it is also further oxidized to acetyl-coenzyme A (CoA) by pyruvate:ferredoxin

oxidoreductase (PFOR). Acetyl-CoA is then converted to acetate by phosphotransacetylase and acetate kinase.

E. proavitum does not ferment ethanol as product (Zheng et al. 2015b) and the alcohol dehydrogenase gene is absent in the genome. However, strain Rs-D17 possesses the bifunctional alcohol/acetaldehyde dehydrogenase gene *adhE*. The deduced AdhE sequence comprised conserved domains of both alcohol dehydrogenase (Adh) and acetaldehyde dehydrogenase (Aldh) (Fig. S4.2). Moreover, one NAD-binding site was also observed, indicating its potential dual functions of Adh and Aldh, which produce ethanol and acetaldehyde from acetyl-CoA (Fig. 4.2) (Koo et al. 2005). The phylogenetic analysis showed the *adhE* gene, which is absent in the ancestral *E. proavitum*, of Rs-D17 was more closely related to those from *Clostridium* species than that of *E. minutum* (Fig. S4.3). This implies that this gene was acquired by horizontal gene transfer.

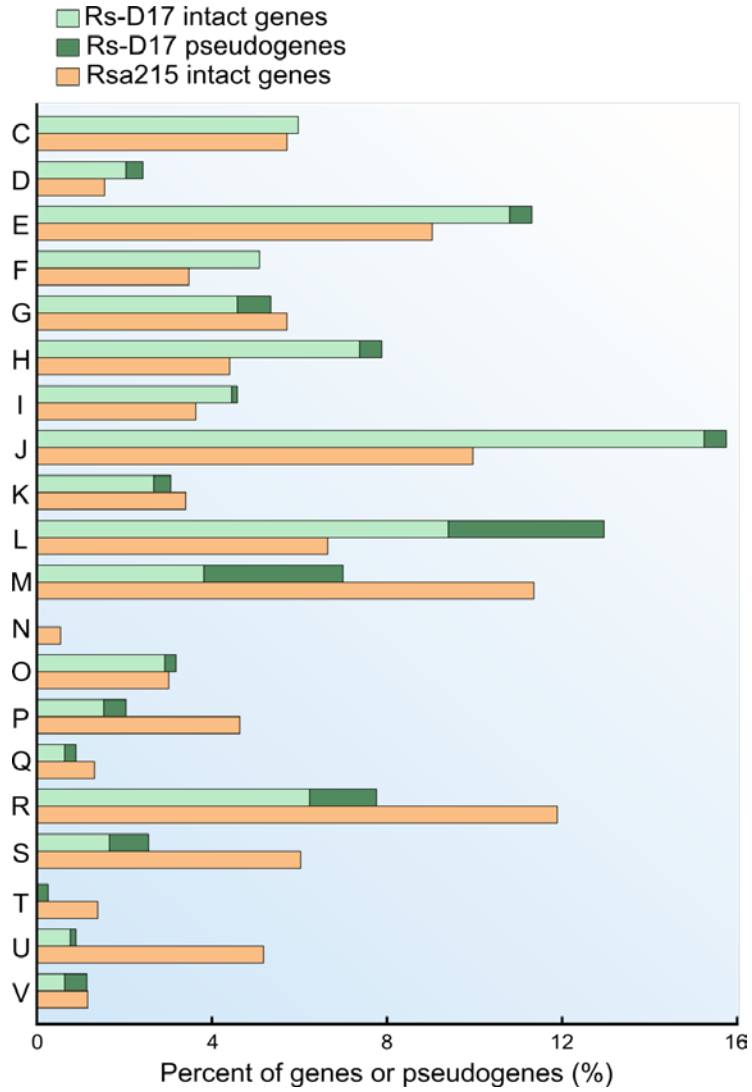

Fig. 4.1. Cluster of orthologous group of proteins (COGs) analysis of intact genes and pseudogenes in the genomes of *E. proavitum* and Rs-D17.

Hydrogen and ethanol fermentation. *E. proavitum* possesses both membrane-bound [NiFe]-hydrogenase and soluble [FeFe]-hydrogenase, potentially involved in hydrogen formation. The stoichiometry of the dissimilatory metabolism shows that one H_2 from NADH and one H_2 from reduced ferredoxin, which corresponds the co-occurrence of the two types of hydrogenase. The [NiFe] hydrogenase operon of *E. proavitum* comprises the genes encoding the typical subunits and maturation proteins, which share synteny

with those from *E. minutum*. Comparative analysis of the genes coding for the large subunit of [NiFe] hydrogenase (*echE*) revealed that the sequences of *E. proavitum* and *E. minutum* are closely related (Fig. S4.4), suggesting the hydrogen fermentation pathway is conserved between these two strains and the [NiFe] hydrogenase may also functions as redox-driven ion pumps (Herlemann et al. 2009), and this type of energy conservation may be present also in *E. proavitum* (Fig. 4.2). The gene of large subunit of [FeFe]-hydrogenase (*hydA*) from *E. proavitum* and strain Rs-D17 are more closely related than *E. minutum* (Fig. S4.5; Zheng et al. 2013), while strain Rs-D17 has only one cluster of [FeFe]-hydrogenase genes.

The reduction of NADH to hydrogen is thermodynamically unfavorable when hydrogen partial pressure is high (Schink 1997), [FeFe]-hydrogenase is probably not involved in hydrogen formation of endosymbionts locating in the cytoplasm of flagellate, where hydrogen accumulates to high concentrations (Pester and Brune 2007) by the hydrogenosomes are located (Lindmark and Muller 1973). Instead, the NADH formed during glycolysis is probably regenerated by the reduction of acetyl-CoA to ethanol. The H_2-rich environment may lead the endosymbionts to alter their reduced fermentation product from H_2 to ethanol (Hongoh et al. 2008) and the [NiFe] hydrogenase was lost by the endosymbionts during the evolution.

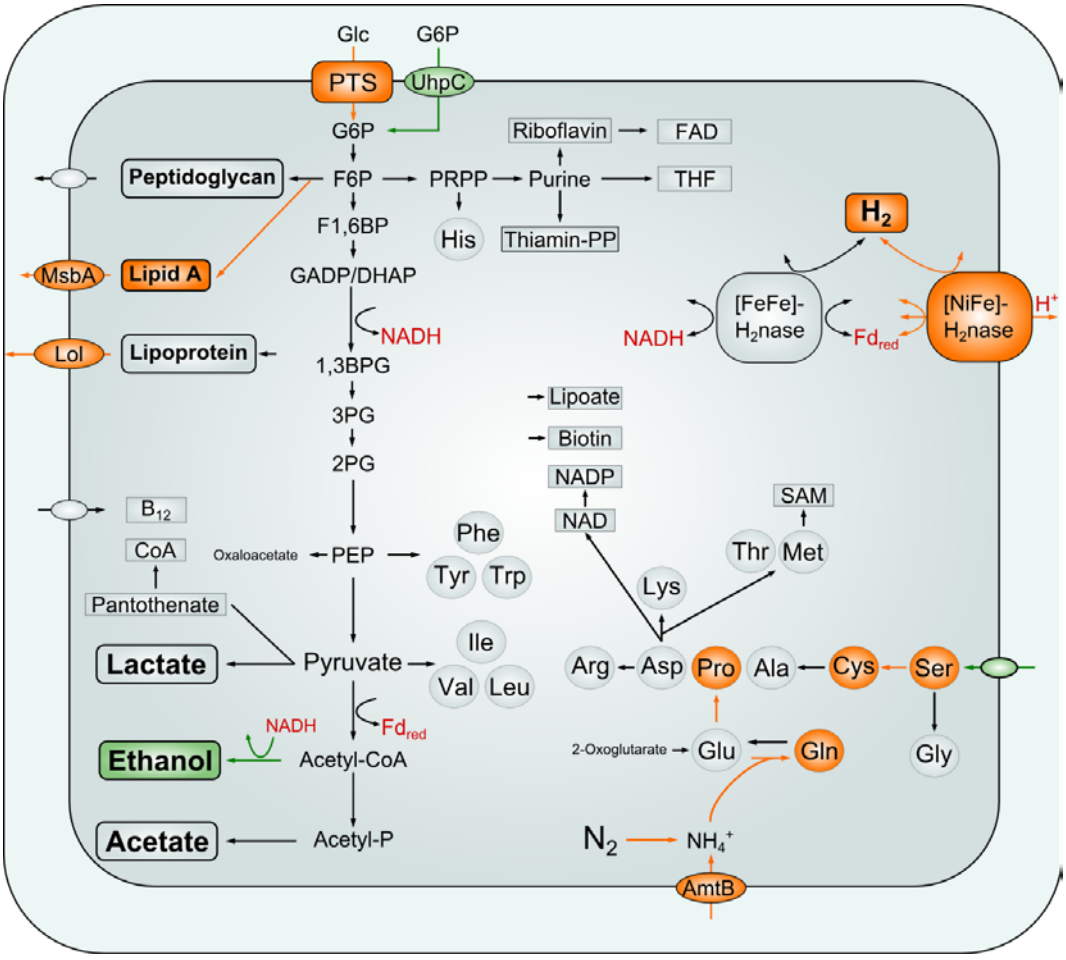

Fig. 4.2. Reconstruction of metabolic pathways in *E. proavitum* in comparison with those of Rs-D17. The pathways and products that are specific for *E. proavitum* are shown in orange, those specific for Rs-D17 are in green, and those shared by both are in grey. Products involving electron transfer are shown in red.

Amino acids, co-factors synthesis. *E. proavitum* can synthesize all essential amino acids (EAAs) and several non-EAAs as well (Fig. 4.2). This result corresponds to that the growth of *E. proavitum* does not require addition of Casamino acids in the medium (Zheng et al. 2015b). Several genes for proline, cysteine, and serine production, which are absent in Rs-D17, are retained in *E. proavitum*. The glutamine synthetase gene *glnA*, which is pseudogenized in Rs-D17, is still intact in the free-living relative. Although several genes involved in

the aromatic amino acid biosynthesis (*aroH*, *hisC*, and *pheA*) and CoA biosynthesis (*coaD*) are duplicated, probably associated with restriction-modification systems (Zheng et al. unpublished data), this is not the case in *E. proavitum*, which indicates that the duplication must happened relatively recently. All the genes for co-factor synthesis present in Rs-D17 also exist in *E. proavitum*.

Nitrogen metabolism. The genome of *E. proavitum* contains only one cluster of genes encoding the structural components of the nitrogenase complex and the *nifHDK* genes all into the Group IV nitrogenases, which was poorly characterized and so-called as "*nif*-like" group. However, it was shown that *E. proavitum* fixes nitrogen with the Group IV nitrogenase (Zheng et al. 2015b). The whole cluster of *nif* genes are absent in Rs-D17, and the pseudogene of ammonium transporter, *amtB*, is intact in *E. proavitum*.

Defense mechanism and DNA repair system

Compared to other intracellular symbionts, Rs-D17 retains quite a few genes for DNA repair system and all these genes are also in *E. proavitum*. Additionally, *nfo* for base excision repair and *recO* for recombination repair, which are absent or pseudogenized, are present in *E. proavitum* (Table S4.2). All the genes for translesion DNA polymerases (TLPs) are absent in both endosymbionts and free-living strain.

Cell wall biosynthesis interruption

The majority of the pseudogenes of Rs-D17 are classified into the COG categories "M" (cell wall/membrane/envelope biogenesis). A whole cluster of genes for lipid A (a key component of lipopolysaccharide) synthesis are disrupted by stop codons. This cluster of genes was all found intact in *E. proavitum* and the genes did not suffer rearrangements (Fig. S4.6). Correspondingly, the lipid A

transporter gene (*msbA*) and the lipoprotein transporter genes (*lolAC*) are also disrupted in Rs-D17. It is interesting that when we compared the lipid A synthesis genes from a metagenomic library of closely related *Endomicrobia* from a different flagellate specie (CET; IMG Project ID: Gi01566; Ikeda-Ohtsubo 2007) we found the genes were also pseudogenized in CET and they showed different patterns of disruption (Fig. S4.6).

Synonymous and non-synonymous divergence in *Endomicrobium*

The close evolutionary relationships between *E. proavitum* and Rs-D17 indicate that the endosymbiosis was relatively recent in origin. Here, we compared the average size of all *E. proavitum* genes with those of *E. proavitum* orthologs that are classified as intact, absent (lost via large deletion), or pseudogenes (disrupted) in the Rs-D17genomes (Fig. S4.7). It was noted that the size of genes classified as pseudogenes were slightly larger than the others and the size of the remaining classes did not vary significantly. From the current data, it is difficult to determine if there are "cryptic" pseudogenes in Rs-D17, since the length of all classes are not significantly different. The reason may be the number of visible pseudogenes in Rs-D17 that have homologs in *E. proavitum* is limited compared to other endosymbionts with large number of pseudogenes (Clayton et al. 2012).

The number of nonsynonymous substitutions per nonsynonymous site (d_N) and the number of synonymous substitutions per synonymous site (d_S) values were calculated for 611 orthologous genes in *E. proavitum* and Rs-D17 (Fig. 4.3). All the dots are evenly distributed around the dashed line, which means that there is no such genes are under relatively higher pressure of selection. This also implies that there may not be "cryptic" pseudogenes as predicted in other symbiosis system. The median d_N/d_S value is 0.11, which is much lower than 1. It

shows that a strong purifying selection for the intact genes in the genome of strain Rs-D17.

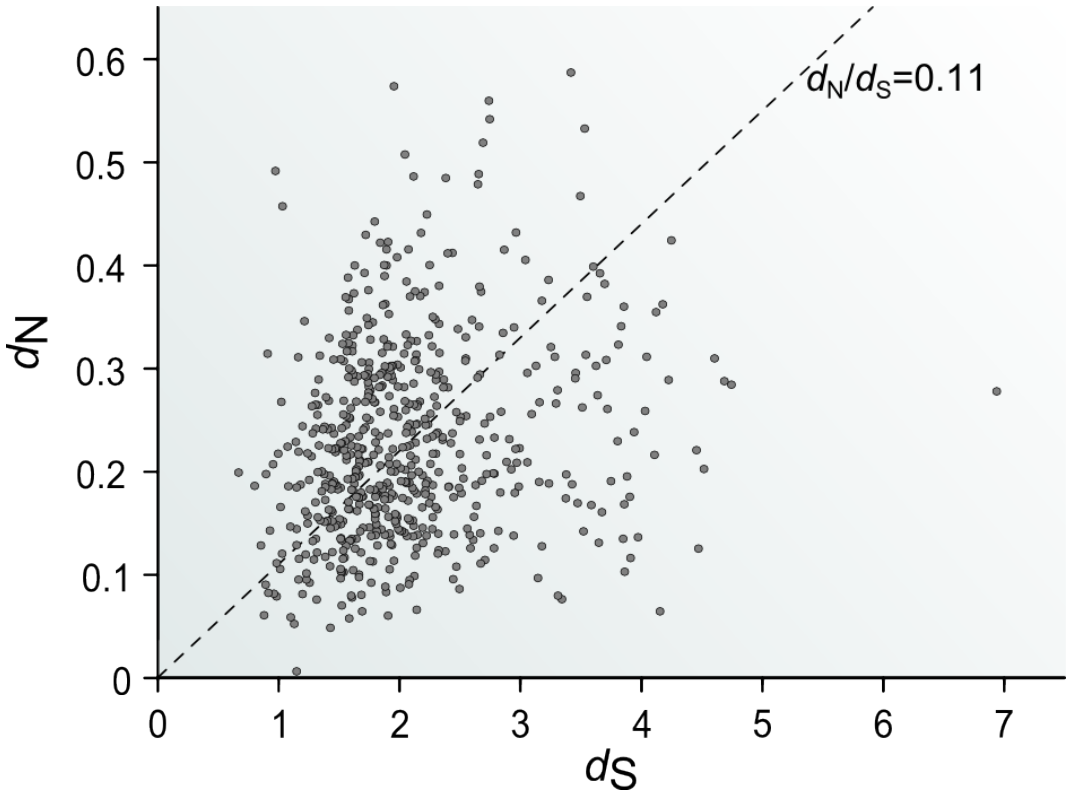

Fig. 4.3. Plot of d_N versus d_S of 611 orthologous intact genes in *E. proavitum* and Rs-D17. Each point represents a single gene. Median d_N/d_S ratio is plotted as a dashed line.

We also calculated the d_N and d_S values of intact orthologous genes in CET to compare with Rs-D17 (Fig. S4.8). It was shown that the Endomicrobium species were not undergoing recombination among different flagellate hosts, in this case, the d_S values for will reflect the divergence time of the symbionts. For the ternary plot of *E. proavitum*, strain Rs-D17, and CET, most genes exhibit similar divergence patterns, with d_S values much lower for RsD17-CET (Fig. 4.4) and the values for CET-Epr and RsD17-Epr are similar, indicating the divergence time of both endosymbionts are similar. We also calculated the d_N/d_S values for both Rs-

D17 and CET. It showed that the values for both genomes are similar, which means the genes are under similar selection power in both endosymbionts. To confirm if the purifying selection power are dependent on the gene functions, we compared the d_N/d_S values for different COG groups of genes in Rs-D17 and CET (Fig. S4.9). The analysis showed that the all the values are similar for different COG groups and they are around the median value (0.1). Similarly, the values for both endosymbionts did not show difference. This means the purifying selection are happening in both genomes of endosymbionts and the selection is function independent.

Discussion

This study focuses on the comparative analyses of the genomes of the termite gut flagellate endosymbiont (strain Rs-D17), and the recently described closely related, free-living strain *E. proavitum*. Before the genome of *E. proavitum* is available, there were only two genomes from the whole phylum of *Elusimicrobia*, the other one, *E. minutum*, is phylogenetically distant and obscure the analysis of the genome reduction of this endosymbiosis relationship. The isolation of *E. proavitum*, the first representative of free-living *Endomicrobia* class, and the completed genome sequence promise a good model system for the study of genome evolution of eukaryotes intracellular symbionts.

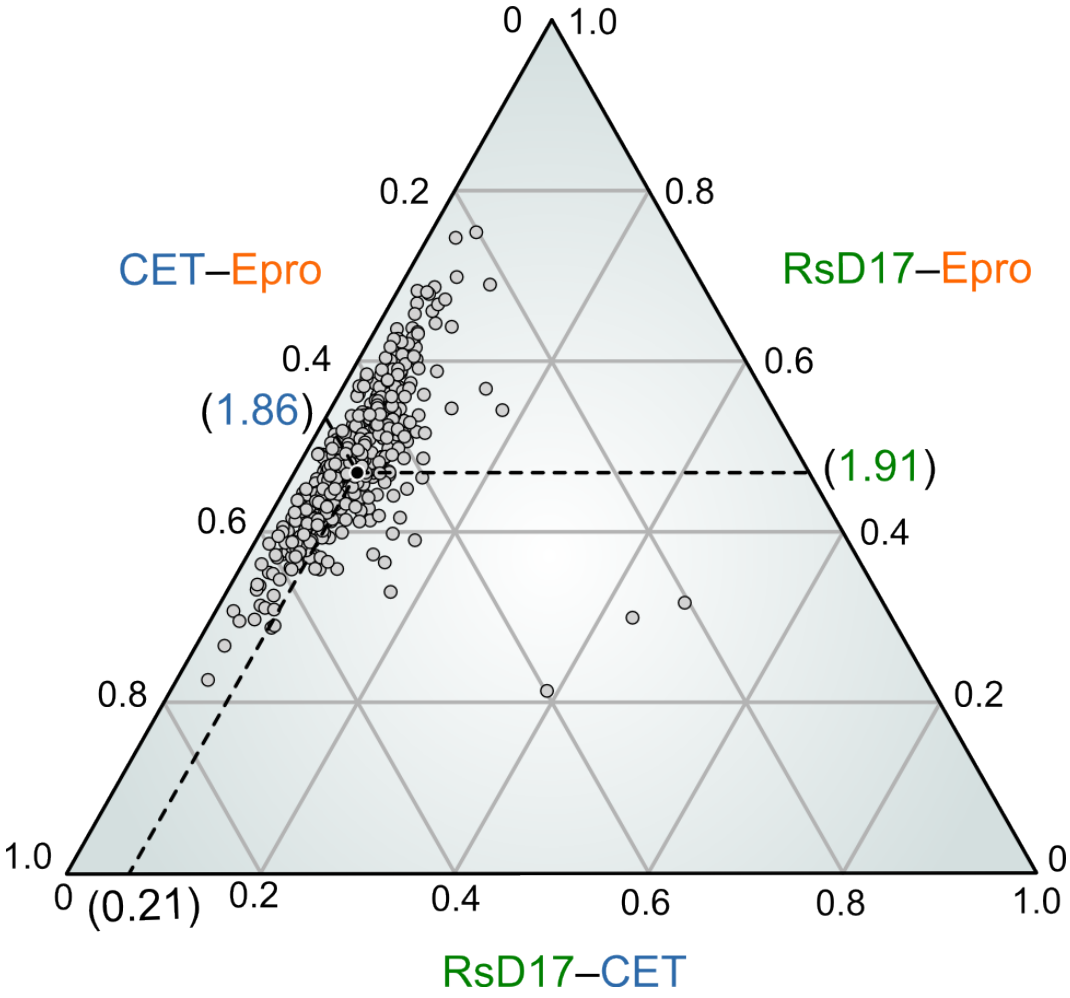

Fig. 4.4. Ternary plots of sequence divergence at synonymous sites for 611 orthologs of *E. proavitum*, strain Rs-D17 and CET. Each dot represents a triplet of orthologs. The sum of all three pairwise d_S values have been normalized to 1 and plotted onto the ternary plot. The absolute median d_S value for each pair is shown on the axes in brackets.

Endosymbionts genome is not only a sub-set of the ancestral genome

It was found that endosymbiont gene repertoires, in most cases, are only a reduced subset from the free-living relative without new genes acquired after the divergence of symbiotic lineages (Santos-Garcia et al. 2014). However, in the genome of strain Rs-D17, we found several pathways are specific for the endosymbionts (Fig. 4.2). The first one is the importer gene for G6P, which is absent in *E. proavitum*. Since the PTS system for glucose transport is missing in

the endosymbionts genome, the main energy and carbons source must be G6P instead of glucose. Termite gut flagellates can metabolise cellulose and produce acetate, CO_2 and H_2. It is unlikely that there is free hexose present intracellularly in the flagellate. The phylogeny of UhpC transporter showed that its closest relative is from *E. minutum*, while they are quite different and its absence in *E. proavitum* implying it was probably acquired by HGT.

The second one is the alcohol/acetaldehyde dehydrogenase gene (*adhE*) and it seems that the gene is acquired by HGT. It was determined that the H_2 partial pressure is extremely high in the hindgut of lower termites (30–70 kpa) (Pester and Brune 2007). It is unlikely that the [FeFe] hydrogenase in Rs-D17 can produce H_2 under this condition. Instead, it may produce ethanol to help reoxidizing NADH to NAD^+ for the next round of glycolysis. At the same time, Rs-D17 lost the entire [NiFe] hydrogenase system because of its surrounding H_2-rich environment.

E. proavitum can synthesize all the essential amino acids and most nonessential ones, which is also supported by their independent growth of supplemented Casamino acids in the medium (Zheng et al. 2015b). Rs-D17 retains 15 amino acids biosynthetic pathways from its ancestral relative implying their ecological function of amino acids provision to their flagellate host (Hongoh et al. 2008).

Cell envelope and defense mechanism

Many genes that are designated as cell envelope synthesis related group are disrupted in Rs-D17 and the whole cluster of lipid A synthesis genes are found pseudogenized, indicating the endosymbionts do not possess a mature outer membrane. The same situation is also found in another endosymbionts from a different flagellate host (Fig. S4.6). It seems that the symbionts are undergoing the

similar evolutionary pressure for the loss of cell wall synthesis. However, the microscopic observations revealed the symbionts are surrounded by two membranes (Stingl et al. 2005), suggesting that the outer membrane must derive from the host. As the genomes of obligately intracellular bacteria shrink, many genes involved in the production of fatty acids, phospholipids and peptidoglycan are lost (McCutcheon and Moran 2012). It seems that the cell wall synthesis must be lost during the early stage of the endosymbiosis association.

Moran et al. (2009) and McCutcheon and Moran (2012) proposed the steps of genome erosion: the crucial turning point is the loss of DNA repair mechanisms, which brings increased mutation rates. Unlike other endosymbionts, Rs-D17 retains numerous genes for DNA repair systems, which are a subset of *E. proavitum*. So the loss of DNA repair systems seems not to be essential for the genome evolution here. In contrast, both of the genomes lack the DNA polymerases for DNA damage tolerance pathway, which provides a mechanism to tolerate DNA lesions during replication. The *umuC* and *umuD* genes are even not present in the free-living strain indicating this system has been lost before the divergence of free-living and endosymbionts long before. Our results indicate that DNA damage tolerance mechanism is not the main power for the genome evolution either in the free-living or the endosymbiotic strain, while both of them retain many of the DNA repair systems for keeping the current genes. Massive gene inactivation could be driven by the spread of mobile genetic elements (MGEs) (Burke and Moran 2011). Although the endosymbiont suffered a large number of rearrangements (Zheng et al. unpublished data), there are no conventional MGEs in Rs-D17. However, comparative analysis revealed that many regions in the genome that suffered rearrangement were flanked by restriction-modification systems, which occur abundantly in Rs-D17 (Hongoh et

al. 2008). A detailed analysis found that RM systems could behave as MGEs and might drive the erosion by shaping the genomes and accelerate the evolution (Zheng et al. unpublished data).

Purifying selection and parallel evolution

It was shown that the endosymbionts Rs-D17 and its free-living relative *E. proavitum* are quite close by both analyses of 16S rRNA sequence and concatenated ribosomal protein sequences (Zheng et al. 2015b). The two genomes also have many homologous genes between their genomes (Table S4.1), which indicates that the transition of Rs-D17 to endosymbiotic lifestyle is relatively recent. The maximum age of '*Ca*. Endomicrobium trichonymphae' acquired as endosymbiont was estimated at about 70 Mya, long after the flagellates entered the termites (Ikeda-Ohtsubo and Brune 2009). The intracellular community analysis of the host flagellates revealed that the symbionts are vertically transmitted without any recombination, suggesting that the symbionts must experience a reduction in effective population size, which has consequences for their genome evolution (Wernegreen 2002).

In bacterial genomes, pairwise values of d_N/d_S typically range from 0.04 to 0.2 for functional genes that are evolving under stabilizing selection (Rocha et al. 2006). Our analyses revealed that most of the protein coding genes from Rs-D17 have d_N/d_S ratios around 0.1, indicating that the genome of Rs-D17 is undergoing purifying selection. It was shown that in some recently acquired endosymbionts, there are a subset of genes designated as "cryptic pseudogenes" having d_N/d_S ratios greater than 0.3 and relatively smaller gene sizes compared to those having lower d_N/d_S ratios (Clayton et al. 2012). However, Rs-D17 does not have certain group of genes: there are no genes with abnormally high d_N values and the genes classified into different groups do not show significantly different length (Fig.

S4.7). Moreover, we found that the purifying selection do not show any bias, the genes from different COG groups show similar power of selection indicating the genome evolution of Rs-D17 is genome-wise and function independent.

In this study, we also compare the genes from another metagenomic library of *Endomicrobium* symbionts from another flagellate specie. Although the sequenced genome is not circular at the moment, it shows that many genes found pseudogenized in Rs-D17 are also disrupted in CET. But the patterns of pseudogenization are different in both genomes. The cluster genes of lipid A synthesis, as an example, showed the same arrangements in all three genomes, while the genes in the endosymbionts are disrupted by stop codons. However, the positions of the stop codons are different and the genes of endosymbionts show different fragment of sequences deletion compared to that of the free-living strain. It indicates that both of the symbionts are undergoing the selective force to drive inactivation of cell envelope biosynthesis. In addition, the protein-coding genes of both endosymbionts show slightly divergence of d_S values, the d_N/d_S ratios are quite similar indicating that the genomes of endosymbionts from different hosts are undergoing similar purifying selection power and evolutionary rate.

Conclusions

Endomicrobia are a specific lineage of bacteria that are found exclusively abundant in termite gut as the intracellular symbionts of the flagellates. The isolation and complete genome sequencing of the free-living relative, *E. proavitum*, provide a good template for the comparative analysis. It is shown that the endosymbionts could obtain some additional genes to assist the symbiosis association during the evolution. Till now, there is only one complete genome of the endosymbionts is available, more genomes of the endosymbionts and free-

living strains would allow us to understand this multilayer symbiosis system better.

Materials and Methods

Genome sequencing and annotation

Genomic DNA of strain Rsa215 was prepared using cetyltrimethylammonium bromide (CTAB) extraction and commercially sequenced with Pacific Biosciences (PacBio) SMRT sequencing (for details, see Zheng and Brune 2015). Initial annotation was done using the *IMG-ER* suite (Markowitz et al. 2009) and the *MicroScope* platform (Vallenet et al. 2013). The annotations were finalized by manually comparing the annotations obtained from different platforms and checking predicted genes. The distribution of the protein coding genes within the COG (Clusters of Orthologous Groups) functional categories (Tatusov et al. 1997) was determined using the *COGNiTOR* software (Kristensen et al. 2010).

Synteny analyses

The common and specific genes among *Endomicrobia* genomes were determined by *Gene phyloprofile* tool of *MicroScope* platform, using a threshold of ≥ 50% similarity and ≥ 80% of matched length.

Synonymous, non-synonymous analysis

Protein-coding genes were aligned for each pair of *Endomicrobium* genomes and used to estimate divergence at both synonymous (d_S) and non-synonymous (d_N) sites. 624 homologous genes shared among all the three *Endomicrobium* genomes were aligned in frame by generating amino acid alignments with *MUSCLE* (Edgar 2004) and then converting them back into nucleotide sequences. All the columns with gaps were deleted for the following analysis. The

synonymous (d_S) and non-synonymous (d_N) substitution frequency per site was based on maximum likelihood estimation using *codeml* implemented in *PAML* 4.8a (Yang 2007) and the method proposed by Yang and Nielsen (2000). The d_N and/or d_S values were plotted with R packages.

Acknowledgements

This study was funded by the Max Planck Society. Hao Zheng is a doctoral student in the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC), Marburg.

References

Brune A. 2011. Microbial symbioses in the digestive tract of lower termites. In: Rosenberg E ,Gophna U, editors. Beneficial Microorganisms in Multicellular Life Forms. Heidelberg: Springer, pp. 3–25.

Brune A, Dietrich C. 2015. The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. *Annu Rev Microbiol*. doi: 10.1146/annurev-micro-092412-155715.

Burke GR, Moran NA. 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol*. 3:195–208.

Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, Weiss RB, Fisher M, Dale C. 2012. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genet*. 8:e1002990.

de Bary A. 1879. Die Erscheinung der Symbiose: Vortrag gehalten auf der Versammlung Deutscher Naturforscher und Aerzte zu Cassel. Strassburg: Verlag von Trübner KJ.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.

Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P, Brune A. 2009. Genomic analysis of "*Elusimicrobium minutum*," the first cultivated representative of the phylum "*Elusimicrobia*" (formerly Termite Group 1). *Appl Environ Microbiol*. 75:2841–2849.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci U S A*. 105:5555–5560.

Hongoh Y, Ohkuma M. 2010. Termite gut flagellates and their methanogenic and eubacterial symbionts. In: Hackstein JHP, editors. (Endo)Symbiotic Methanogenic Archaea. Heidelberg: Springer, pp. 55–79.

Ikeda-Ohtsubo W. 2007. Endomicrobia in termite guts: symbionts within a symbiont. Doctoral thesis, Philipps-Universität Marburg, Marburg, Germany.

Ikeda-Ohtsubo W, Brune A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol Ecol*. 18:332–342.

Koo OK, Jeong DW, Lee JM, Kim MJ, Lee JH, Chang HC, Kim JH, Lee HJ. 2005. Cloning and characterization of the bifunctional alcohol/acetaldehyde dehydrogenase gene (*adhE*) in *Leuconostoc mesenteroides* isolated from kimchi. *Biotechnol Lett*. 27:505–510.

Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin EV, Mushegian A. 2010. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics*. 26:1481–1487.

Lindmark DG, Muller M. 1973. Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. *J Biol Chem*. 248:7724–7728.

Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics*. 25:2271–2278.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 10:13–26.

Moran NA, McLaughlin HJ, Sorek R. 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*. 323:379–382.

Moran NA, Bennett GM. 2014. The tiniest tiny genomes. *Annu Rev Microbiol*. 68:195–215.

Pester M, Brune A. 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J*. 1:551–565.

Rocha EPC, Smith JM, Hurst LD, Holden MTG, Cooper JE, Smith NH, Feil EJ. 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol*. 239:226–235.

Santos-Garcia D, Latorre A, Moya A, Gibbs G, Hartung V, Dettner K, Kuechler SM, Silva FJ. 2014. Small but powerful, the primary endosymbiont of moss bugs, *Candidatus* Evansia muelleri, holds a reduced genome with large biosynthetic capabilities. *Genome Biol Evol*. 6:1875–1893.

Schink B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Mol Biol Rev*. 61:262–280.

Stingl U, Radek R, Yang H, Brune A. 2005. "*Endomicrobia*": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol*. 71:1473–1479.

Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science*. 278:631–637.

Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fevre F, Longin C, Mornico D, Roche D, et al. 2013. MicroScope—an integrated

microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res*. 41:D636–D647.

Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet*. 3:850–861.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 17:32–43.

Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.

Zheng H, Bodington D, Zhang C, Miyanaga K, Tanji Y, Hongoh Y, Xing XH. 2013. Comprehensive phylogenetic diversity of [FeFe]-hydrogenase genes in termite gut microbiota. *Microbes Environ*. 28:491–494.

Zheng H, Dietrich C, Thompson CL, Meuser K, Brune A. 2015a. Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microbes Environ*. 30:92–98.

Zheng H, Dietrich C, Radek R, Brune A. 2015b. *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. *Environ Microbiol*. doi: 10.1111/1462-2920.12960.

Zheng H, Brune A. 2015. Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum *Elusimicrobia*). *Genome Announc*. 3:e00679-15.

Supplementary Information

Loss and gain: Genome evolution of termite gut free-living and flagellate endosymbiontc *Endomicrobia*

Hao Zheng, Carsten Dietrich, Andreas Brune

Supplementary tables

Table S4.1. Gene annotation list of *E. proavitum*.

(https://www.dropbox.com/s/po6um80mobx2317/Gene_annotation.xlsx?dl=0)

Table S4.2. Genes involved in DNA repair and recombination of *E. proavitum*, strain Rs-D17, and some other intracellular symbionts.

(https://www.dropbox.com/s/q297fl7hggwpa1b/Repair_and_recombination_genes.xlsx?dl=0)

Supplementary figures

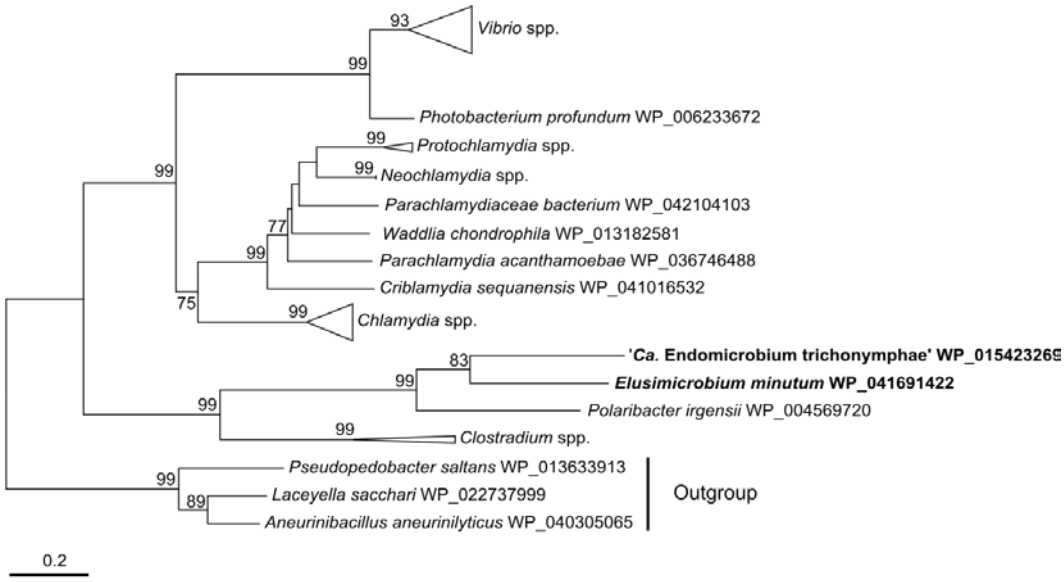

Fig. S4.1. Maximum-likelihood tree of glucose-6-phosphate transporter gene *uhpC* of '*Ca*. Endomicrobium trichonymphae' strain Rs-D17, *Elusimicrobium minutum* (in bold) and other bacterial species. The trees are based on deduced amino acid sequences and were rooted using the sn-glycerol-3-phosphate transporter gene. A total of 1,999 amino acid sites were used to construct the tree with the selected best model (Dayhoff+G+I+F).

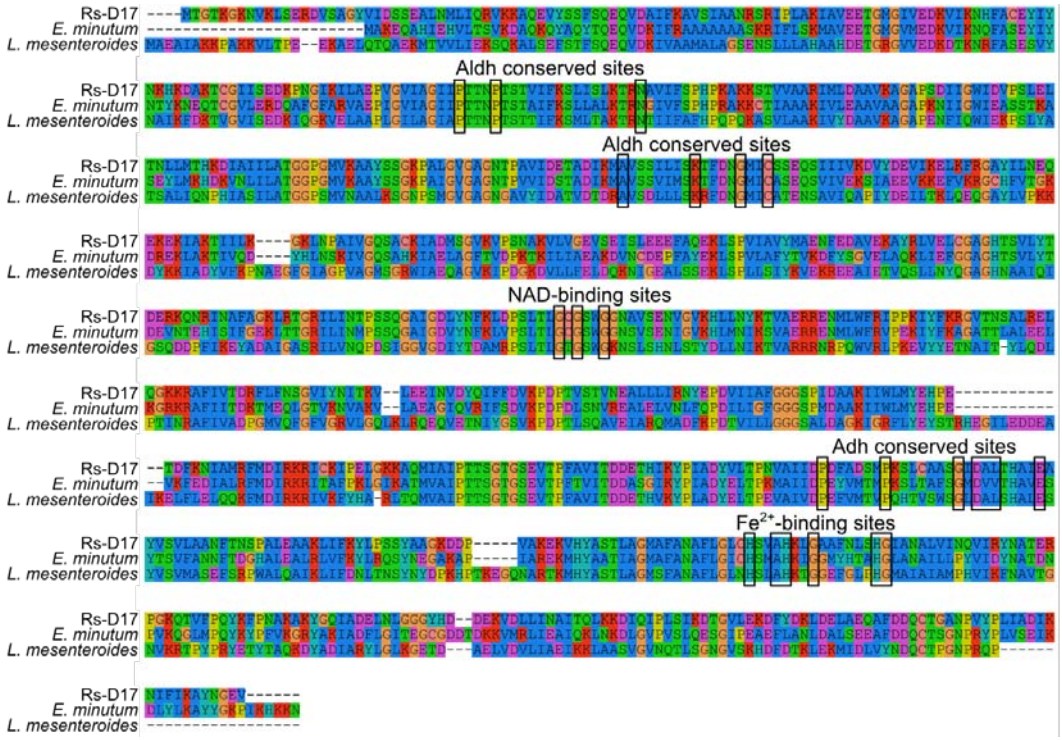

Fig. S4.2. Alignment of deduced alcohol/aldehyde dehydrogenase of '*Ca*. Endomicrobium trichonymphae', *Elusimicrobium minutum*, and *Leuconostoc mesenteroides* C7. The conserved domains of aldehyde dehydrogenase, alcohol dehydrogenase, NAD-binding sites, and Fe^{2+}-binding sites are marked by black boxes based on the motifs described previously (Koo et al. 2005). The sequences were aligned using *MAFFT* and visualized by *SeaView* platform (Gouy et al. 2010).

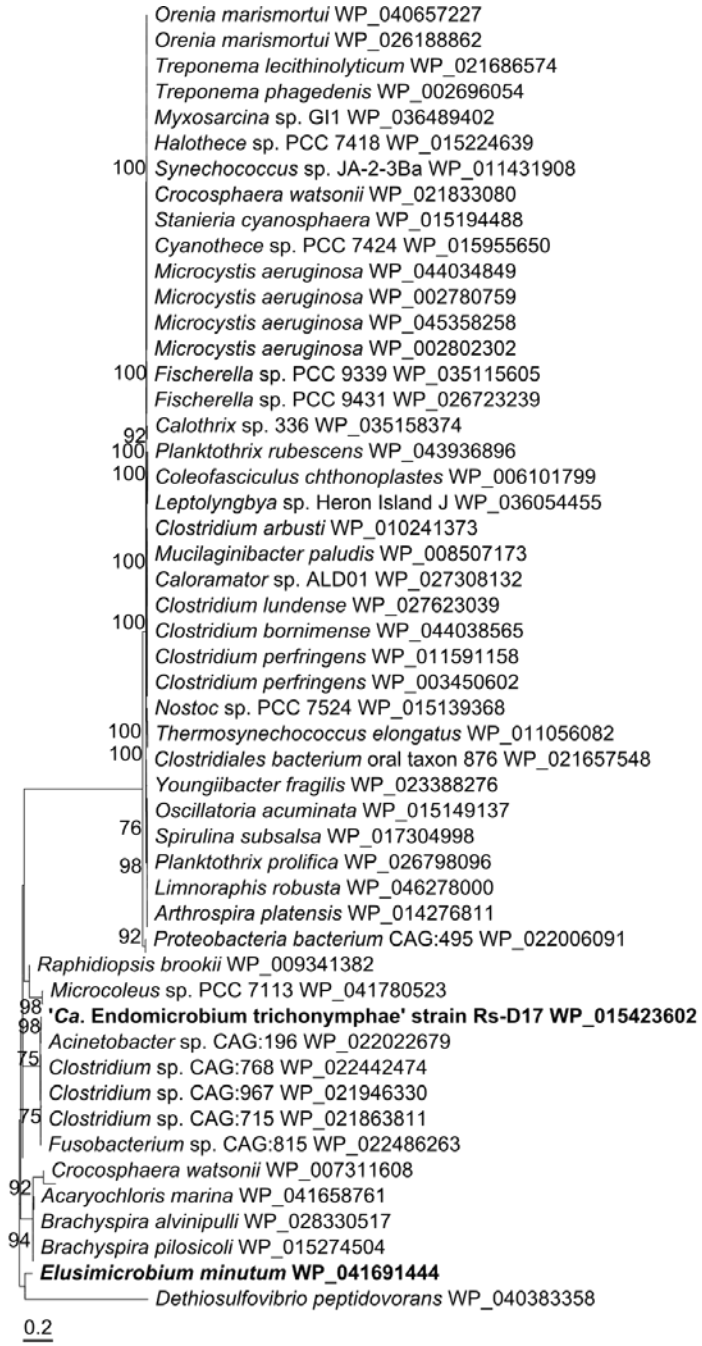

Fig. S4.3. Phylogenetic positions of aldehyde-alcohol dehydrogenase of '*Ca*. Endomicrobium trichonymphae', *E. minutum*, and other bacterial species. Deduced protein sequences were aligned using the MAFFT program (Katoh and Standley 2013), 881 sites were used for the analysis. The maximum likelihood tree was inferred using the MEGA6 program (Tamura et al. 2013) with the best selected model (WAG+G+F). Numbers at nodes indicate percent bootstrap values when greater than 75%.

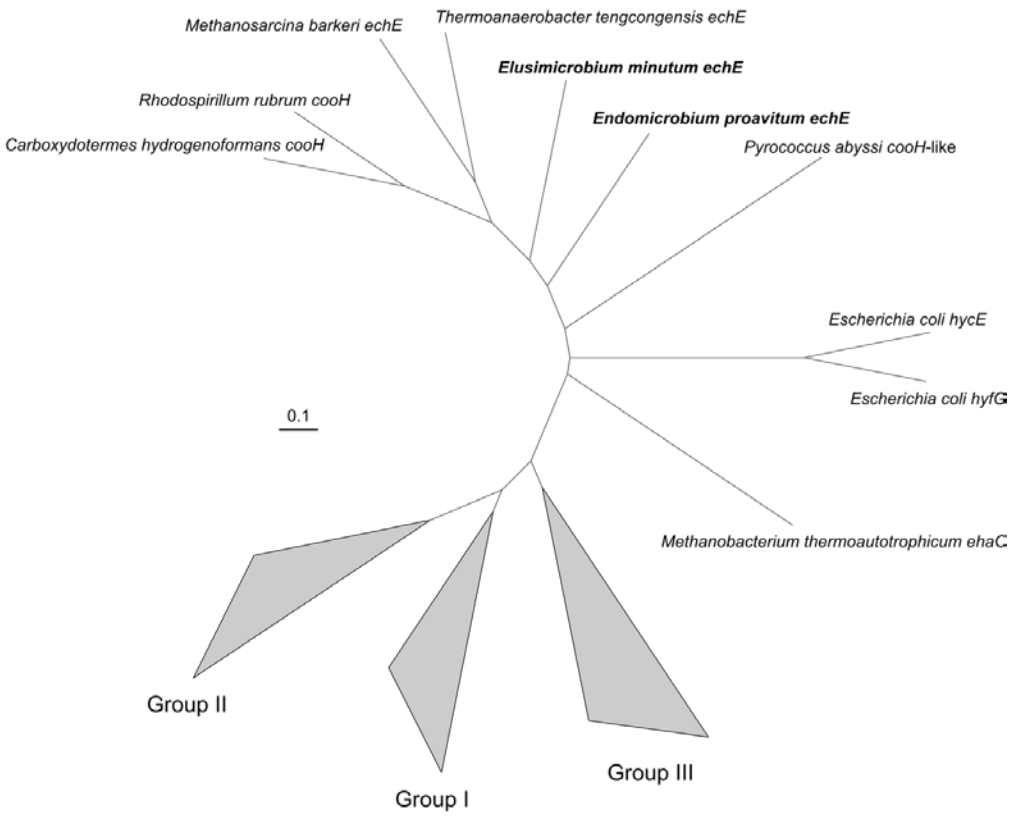

Fig. S4.4. Maximum-likelihood tree of [NiFe]-hydrogenases regenerated from that published previously (Herlemann et al. 2009), based on the deduced amino acid sequences of the large subunit. Both sequences of *E. proavitum* and *E. minutum* fall within the radiation of the sequences assigned to group IV [NiFe] hydrogenases.

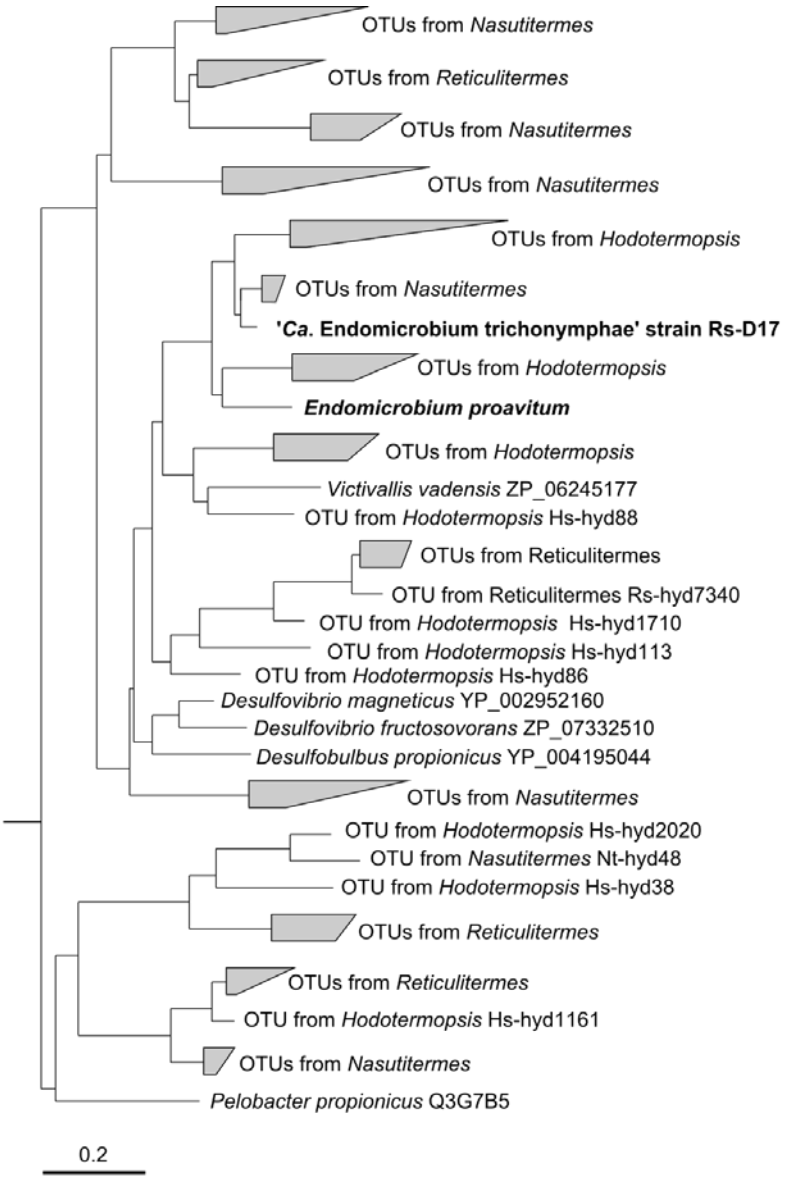

Fig. S4.5. Maximum-likelihood tree of [FeFe]-hydrogenase based on deduced amino acid sequences of the large subunit gene *hydA*. Recalculated from Zheng et al. (2013).

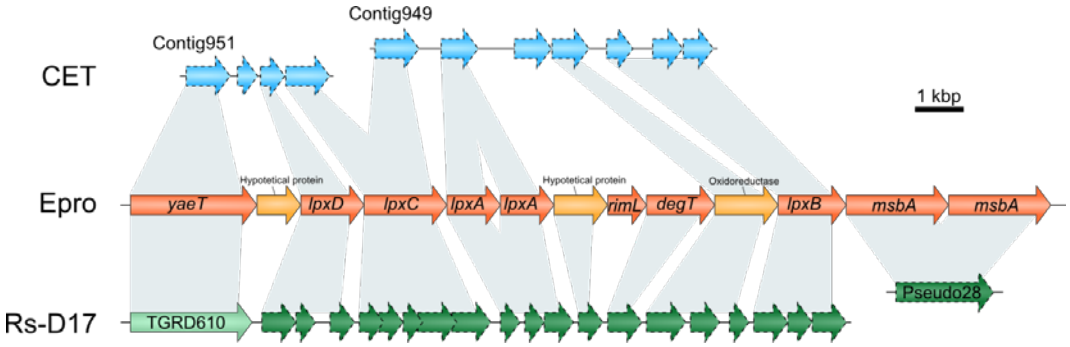

Fig. S4.6. Lipid A biosynthesis related genes cluster of CET (blue), *E. proavitum* (orange) and strain Rs-D17 (green). Pseudogenes are shown in dashed arrows.

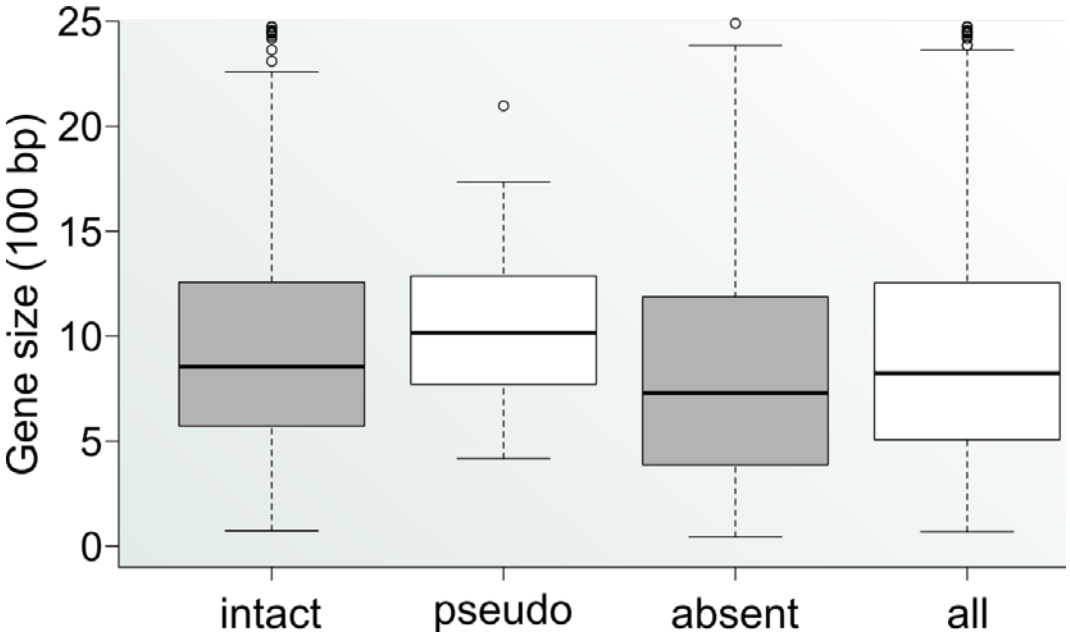

Fig. S4.7. Box plots of the sizes of *E. proavitum* genes that are still intact, pseudogenized (pseudo), or absent in Rs-D17, and all *E. proavitum* intact genes (all).

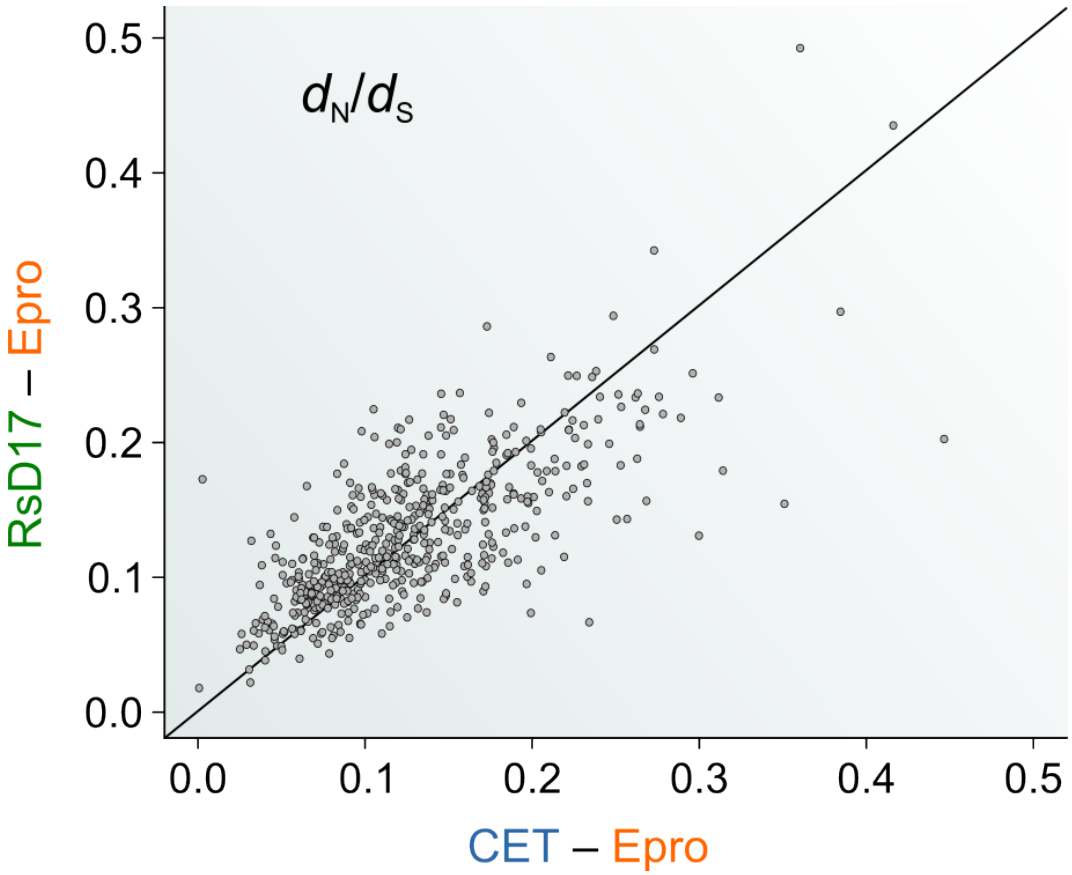

Fig. S4.8. Plot of Rs-D17 dN/dS versus CET dN/dS for 611 orthologous genes in *E. proavitum*, Rs-D17, and CET. Each point represents a single intact orthologous gene.

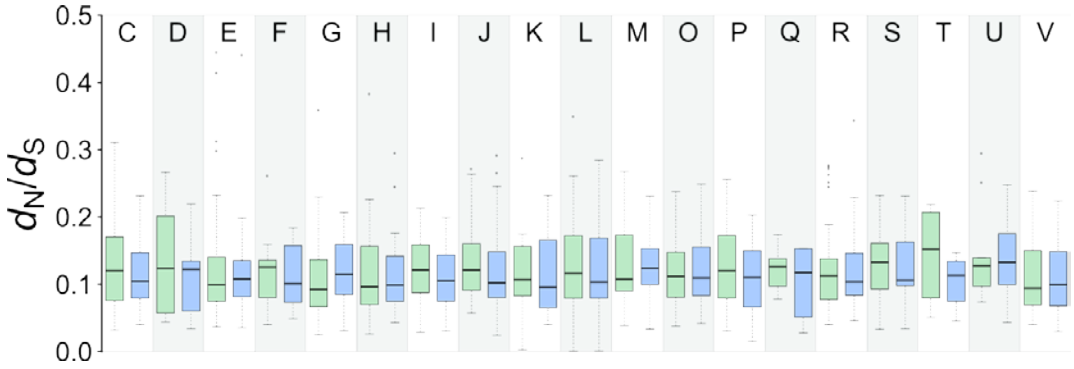

Fig. S4.9. d_N/d_S values of orthologous genes between *E. proavitum* vs. strain rs-D17 (green) and *E. proavitum* vs. CET (blue) ,which belong to different COGs.

Chapter 5

Restriction-modification systems as mobile genetic elements in the evolution of an intracellular symbiont

Published online in Molecular Biology and Evolution (2015)

doi: 10.1093/molbev/msv264

Hao Zheng[1], Carsten Dietrich[1], Yuichi Hongoh[2], Andreas Brune[1]

[1]Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany; [2]Department of Biological Sciences, Tokyo Institute of Technology, Tokyo, Japan.

Contributions: H.Z. planned and carried out experiments, analyzed data, discussed results, and wrote the manuscript. C.D. analyzed data, discussed results. Y.H. discussed results. A.B. discussed results, wrote the manuscript, and secured funding.

Abstract

Long-term vertical transmission of intracellular bacteria causes massive genomic erosion and results in extremely small genomes, particularly in ancient symbionts. Genome reduction is typically preceded by the accumulation of pseudogenes and proliferation of mobile genetic elements, which are responsible for chromosome rearrangements during the initial stage of endosymbiosis. We compared the genomes of an endosymbiont of termite gut flagellates, '*Candidatus* Endomicrobium trichonymphae', and its free-living relative *Endomicrobium proavitum* and discovered many remnants of restriction-modification (R-M) systems that are consistently associated with genome rearrangements in the endosymbiont genome. The rearrangements include apparent insertions, transpositions, and the duplication of a genomic region; there was no evidence of transposon structures or other mobile elements. Our study reveals a so far unrecognized mechanism for genome rearrangements in intracellular symbionts and sheds new light on the general role of R-M systems in genome evolution.

Introduction

Mutualistic symbioses between bacteria and eukaryotes are ubiquitous in nature, and they can greatly affect the genomes of bacteria (Bennett and Moran 2015). Intracellular bacterial symbionts are notable among cellular life forms because of their extremely reduced genomes, which allow exploration of the evolutionary processes that occur during the transition from free-living to endosymbiotic lifestyles (Moran et al. 2008). The highly reduced genomes of many ancient endosymbionts have been sequenced, but only few younger endosymbionts have been studied, mostly because of the lack of suitable models (Moran and Bennett 2014). This leaves enormous gaps in our knowledge of the forces shaping genome evolution in the early stages of an endosymbiotic association. Nevertheless, it is generally accepted that the early stage of genome erosion in endosymbionts is characterized by proliferation of mobile genetic elements (MGEs), specifically transposons, which promote genome rearrangements that lead to the accumulation of disrupted genes and the first deletions of larger gene regions (McCutcheon and Moran 2012).

Endomicrobia are a class of bacteria (phylum *Elusimicrobia*) that comprise numerous intracellular symbionts of termite gut flagellates (e.g., Stingl et al. 2005; Ohkuma et al. 2007). One of them is '*Candidatus* Endomicrobium trichonymphae', which was acquired from free-living ancestors about 40 to 70 million years ago and since then has been vertically transmitted, leading to cospeciation with its flagellate host (Ikeda-Ohtsubo and Brune 2009; Zheng et al. 2015a). The small genome size of '*Ca*. Endomicrobium trichonymphae' strain Rs-D17 (1.13 Mbp) and the presence of many pseudogenes indicate that it is still in an early stage of genome reduction (Hongoh et al. 2008).

The recent isolation and genome sequencing of *Endomicrobium proavitum*, a close but free-living relative of '*Ca.* Endomicrobium trichonymphae' (Zheng et al. 2015b; Zheng and Brune 2015), provides a unique model for studying the early stages of evolution of an intracellular symbiont. Here, we report evidence that the genome rearrangements in strain Rs-D17 are associated with restriction modification (R-M) systems that apparently act as MGEs – a finding that adds an entirely new facet to the mechanisms involved in genome evolution in endosymbiotic associations.

Results and Discussion

Comparative genome analysis revealed that 82% of the protein-coding genes of strain Rs-D17 have homologs in *E. proavitum*. However, despite a moderate level of average amino acid identity (61.3% ± 0.6%) encoded by the two genomes, a genome-wide alignment revealed that the genome of the endosymbiont underwent a large number of rearrangements (Fig. 5.1); *Mauve* software identified 166 synteny breaks (Fig. S5.1). It was therefore quite unexpected that we could not find any conventional MGEs, such as prophage DNA or transposons. Instead, the genome of the endosymbiont contains many R-M systems (Fig. 5.1; Table S5.1), which are scarce in *E. proavitum* (Table S5.2).

R-M systems are composed of genes encoding restriction endonucleases (R) and methyltransferases (M) and serve to cleave any DNA that has not been modified appropriately (Pingoud 2004). It is widely believed that R-M systems are maintained by bacteria to defend the cells from viral infection and the entry of other foreign DNA (Kobayashi 2001). Strain Rs-D17 contains 18 R-M systems and 5 solitary M genes (Table 1), which exclusively belong to Type I–III. Most of the R-M genes belong to Type II, the most common type in prokaryotes (Oliveira et al. 2014). Most R-M systems in strain Rs-D17 have been inactivated by

pseudogenization of one or more genes (Table 5.1). A truncation of R-M genes has been reported also for other bacteria (Furuta et al. 2014).

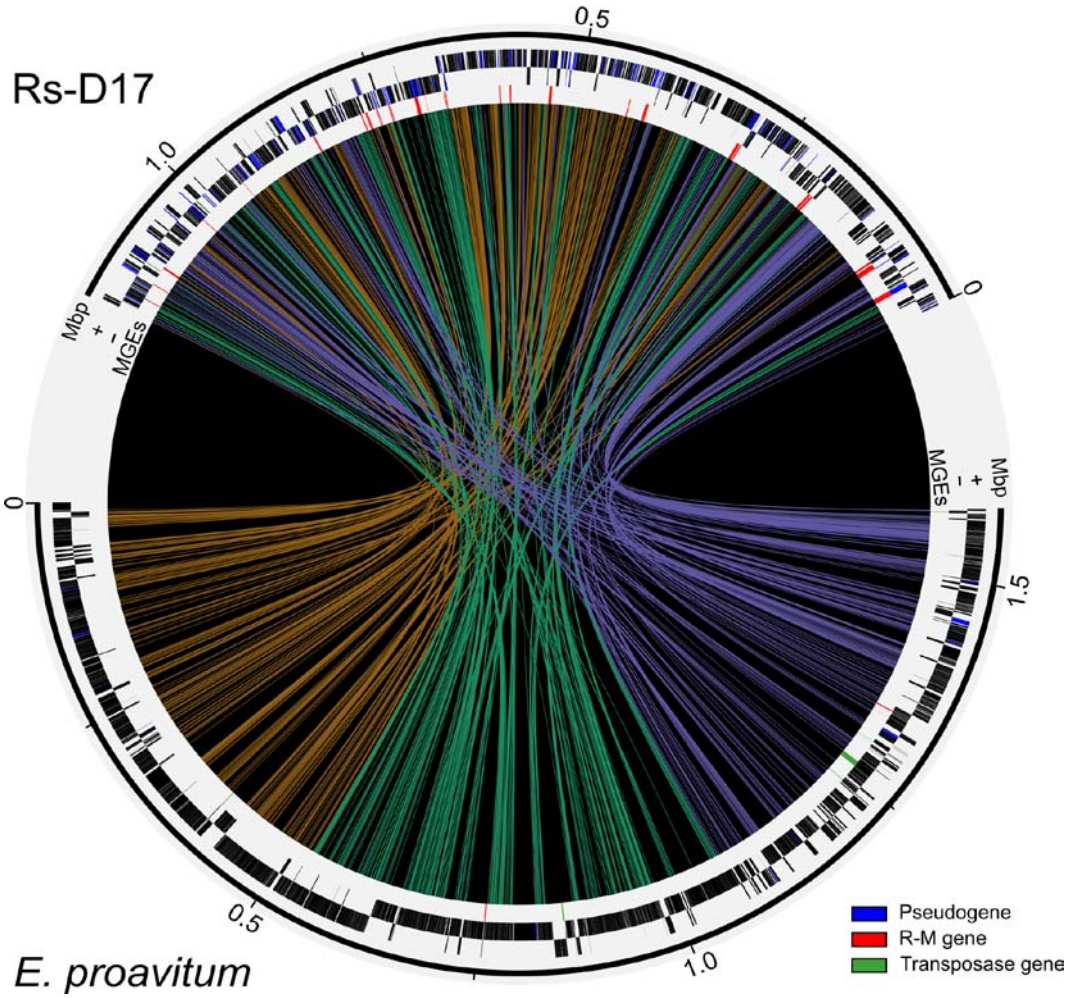

Fig.5.1. Comparison of genomes of '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17 (upper half) and its closest free-living relative *Endomicrobium proavitum* (lower half). The concentric rings denote the following features (from outside): Number of base pairs; ORFs on the forward (+) and reverse (–) strands; and mobile genetic elements (MGEs). In the innermost graph, the homologous genes in the two genomes are connected by lines. A GC skew diagram of the two genomes is shown in Fig. S5.2.

While none of the R-M genes of strain Rs-D17 have homologs in the genome of the ancestral *E. proavitum*, most of them were found also in a metagenomic library of closely related *Endomicrobia* from a different flagellate species (Table S5.1; Zheng et al. unpublished data), which suggests that the R-M genes were

Table 5.1. Restriction-modification (R-M) systems and solitary methyltransferase genes in '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17. See Table S5.1 for a full list of R-M genes in the genome.

R-M systems	Number					Rearrangement patterns		
	R	M	S	Genes (pseudo)	R-M systems (inactive)	Insertions	Duplications	Transpositions
Type I	2	2	2	6 (2)	2 (1)	0	0	2
Type II[a]	11	17	n.a.	28 (25)	13 (13)	7	1	5
Type III	4	3	n.a.	7 (5)	3 (2)	1	0	2
Solitary M genes[b]	n.a.	5	n.a.	5 (5)	n.a.	4	0	1
Total	17	27	2	46 (37)	18 (16)	12	1	10

R, restriction endonuclease; M, methyltransferase; S, DNA sequence specificity; pseudo, pseudogenes; inactive, inactive R-M systems with R or M pseudogenes; n.a., not applicable.

[a]One system corresponds to Type IIC (see Table S5.1).

[b]All solitary M genes were classified as Type II methyltransferases. They were considered solitary if they were at least 10 genes away from the next R-M system (Oliveira et al. 2014) or if they recognized different sequences (i.e., the neighboring M genes B3 and B4 in Table S5.1).

acquired after the separation of free-living and endosymbiotic lines. The top BLASTP hits of the R-M genes are associated with diverse and only distantly related bacterial taxa (Table S1), which agrees with the concept that R-M systems have undergone frequent horizontal gene transfer (Rocha et al. 1999; Furuta and Kobayashi 2013). A comparison of pentanucleotide frequencies revealed that the Euclidean distances between certain R-M genes and other protein-coding genes in strain Rs-D17 are significantly larger than the average intragenomic distance (Fig. S5.3; supplementary Materials and Methods); this also indicates horizontal gene transfers. However, movement of target recognition domains among R-M systems (Furuta and Kobayashi 2012) cannot be evaluated since most of the genes are pseudogenized.

R-M systems often co-occur with CRISPR-Cas systems, which sometimes act synergistically in the defense against foreign DNA (Dupuis et al. 2013). The genome of strain Rs-D17 contains two CRISPR-Cas systems (Type I-C and Type II-C), which comprise the typical arrays of direct repeats separated by short spacer sequences close to the *cas* genes (Fig. S5.4A). They are present also in *E. proavitum*, and in the case of the Type II-C system, also in *Elusimicrobium minutum* (Herlemann et al. 2009), which indicated that the latter system may be conserved in this phylum. In addition, the genome of strain Rs-D17 contains a third *cas* gene cassette (Type II-C) without a CRISPR locus (Fig. S5.4B). Since R-M and CRISPR-Cas systems do not neighbor each other, it is unlikely that their co-occurrence in the genome of strain Rs-D17 is due to a co-transfer in “defense islands” (Oliveira et al. 2014). Although the *cas* genes and the repeat sequences are conserved between strain Rs-D17 and *E. proavitum* (Fig. S5.5), they do not share the same spacer sequences (Table S5.3), which indicates that the two organisms have experienced entirely different invasion situations. Since the spacer

sequences of the two CRISPR loci do not match other regions of the genome, the CRISPR-Cas systems are unlikely to cleave endogenous DNA (Cong et al. 2013).

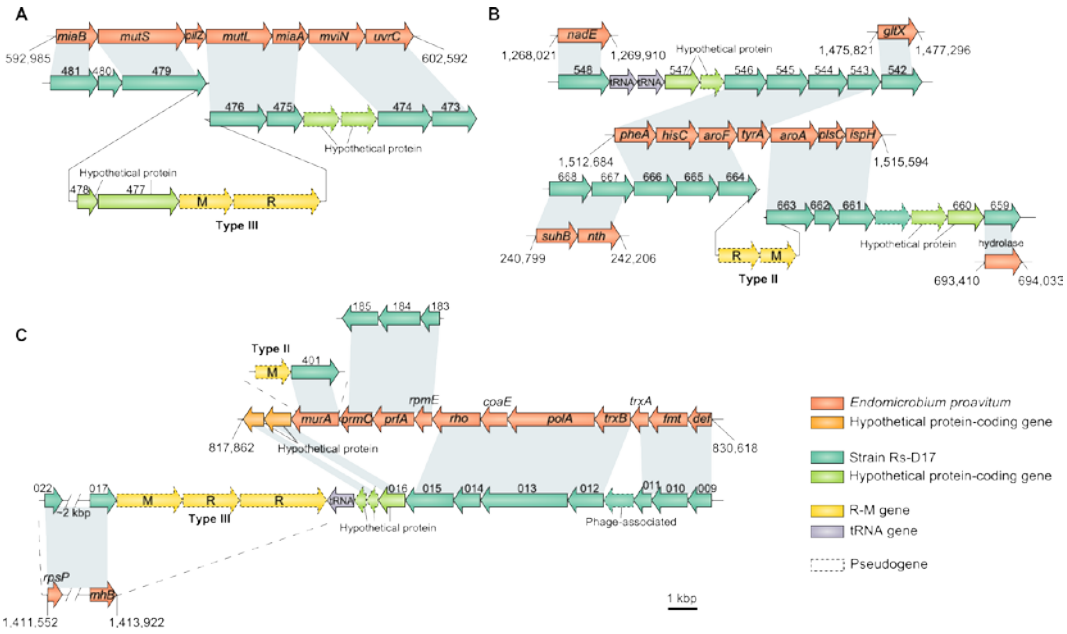

Fig. 5.2. Examples of genome rearrangement sites flanked by restriction-modification (R-M) systems in '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17. (A) Simple insertion, here accompanied by foreign genes. (B) Region duplication. (C) Complex rearrangement, involving several transpositions and inversions. The number above each gene of strain Rs-D17 is the locus tag used in the annotation list of Hongoh et al. (2008). The positions of start and end points of the gene clusters in *E. proavitum* are indicated (numbers). A detailed analysis of all rearrangement sites in the genome of strain Rs-D17 that are associated with R-M systems is shown in Fig. S5.6. The frequencies of these events are shown in Table 5.1.

When we compared the genome regions flanking R-M genes in strain Rs-D17 to the syntenic regions in *E. proavitum*, we found that they were consistently associated with sites of obvious genome rearrangements. The most common rearrangements were insertions and transpositions; a duplication of a genomic region was observed only once (Table 5.1). A typical example of an insertion is shown in Fig. 5.2A, where a pseudogenized R-M gene set, together with two heterologous genes encoding hypothetical proteins, interrupts the space between two DNA mismatch repair (MMR) genes (*mutS* and *mutL*), which are contiguous

in *E. proavitum*. This insertion is of particular interest in an intracellular symbiont because it might suppress the MMR pathway and thereby reduce the stringency of homologous recombination (Li 2008), causing gene deletions and, subsequently, a decreased genome size. The second example (Fig. 5.2B) shows a genomic region that is contiguous in *E. proavitum* but interrupted in strain Rs-D17 by a pseudogenized M gene. During its insertion between the *aroF* and *aroA* genes the pseudogene apparently replaced the *tyrA* gene. Interestingly, one of the flanking regions, including *tyrA*, was duplicated in this process, which possibly rescued the capacity of the endosymbiont for biosynthesis of aromatic amino acids and their provision to the host (Hongoh et al. 2008). In contrast to the situation in *Helicobacter pylori* (Furuta et al. 2011), this duplication was not associated with inversions. The third example shows a complex series of transpositions (Fig. 5.2C) that led to a major rearrangement of a larger genome region: several genes (*rpmE–prmC*; *murA*) contiguous in *E. proavitum* have been separated and moved to different regions in strain Rs-D17 (the *murA* gene is now flanked by an M gene). In addition, a larger stretch of DNA (ca. 2 kb) located distally from that region in *E. proavitum* has been inserted proximally to the original site of transposition and is now associated with a disrupted R-M system. While the consequences of this rearrangement remain unclear, other rearrangements should directly affect the metabolic capacities of the endosymbiont. For example, the pseudogenization of the *glnA* gene encoding glutamine synthetase, which limits the ability of strain Rs-D17 to incorporate ammonia and requires the import of glutamine (Hongoh et al. 2008), may be related to the transposition event responsible for the presence of the flanking R-M system (A17 in Fig. S5.6).

Previous studies comparing the genomes of closely related prokaryotes have already documented that R-M systems are associated with different types of

genome rearrangements (Tsuru et al. 2006; Furuta and Kobayashi 2013). The present study is the first case where R-M systems acting as MGEs are responsible for the genome rearrangements characteristic of the early stage of genome reduction in an intracellular symbiont. It has been shown that an apparent mobility of R-M systems can be caused by their carriage on other mobile elements or by forming a composite transposon with other insertion sequences (Furuta et al. 2010; Takahashi et al. 2011). However, this is unlikely in the case of strain Rs-D17. Despite intensive searches (see supplementary Materials and Methods), we found no evidence for the presence of other MGEs. Moreover, the R-M genes of strain Rs-D17 were never flanked by direct repeats, which are involved in the multiplication of R-M systems in other bacteria (Nobusato et al. 2000; Sadykov et al. 2003). In view of the relatively young age of the symbionts, it is unlikely that mobile elements were originally present but eventually lost from the genomes. However, it is possible that R-M systems themselves act as MGEs (Furuta and Kobayashi 2013).

Genome reduction in endosymbionts depends on the age of the association and is considered to occur in several stages (McCutcheon and Moran 2012). During the initial transition to an intracellular lifestyle, many genes of the endosymbionts are no longer required and the reduced strength of purifying selection leads to rapid accumulation of pseudogenes and rearrangements throughout the genome. In older endosymbionts, the ongoing deletion removes pseudogenes and mobile elements, resulting in highly compact genomes (Moran and Bennett 2014). The relatively recent acquisition of intracellular *Endomicrobia* (Ikeda-Ohtsubo and Brune 2009) matches both the strong accumulation of pseudogenes and massive genome rearrangements. The intermediate genome size of strain Rs-D17 is 29% smaller than that of its free-living relative (1.59 Mbp;

Zheng and Brune 2015), probably due to its strictly vertical transmission, but still substantially larger than the genomes of the more ancient endosymbionts, which can be extremely reduced (Moran and Bennett 2014).

So far, the only model for a nascent symbiosis that allows a comparison of the genomes of both intracellular symbionts and their free-living relative is the *Sodalis*-allied clade of insect endosymbionts and the closely related isolate, strain HS (Clayton et al. 2012). In this case, the primary endosymbionts of insects show numerous genomic rearrangements that are attributed mostly to insertion sequence (IS) elements that are abundant in the genome (Oakeson et al. 2014). Although the massive genome rearrangements in *Endomicrobia* endosymbionts seem to be caused by a different mechanism (i.e., involving R-M systems as MGEs), they led to the same result. Apparently, the relaxed selection in endosymbionts generates many dispensable genes ("cryptic" pseudogenes; Clayton et al. 2012) that become a landing place for any MGEs.

The reasons for the presence of R-M systems but not IS elements in the genomes of endosymbiotic *Endomicrobia* remain unclear. Since the gut flagellates phagocytose wood particles and extracellular bacteria (Yamaoka and Nagatani 1977), it is possible that phages from the environment can invade also the intracellular bacteria located in the cytoplasm of the host cell. In that case, the presence of R-M systems would provide a competitive advantage in the establishment of a stable symbiosis. Alternatively, it is possible that the R-M systems accelerate the evolution of endosymbionts. It has been shown that DNA methylation can drive adaptive genome evolution: changes in the methylome lead to changes in the transcriptome and eventually in phenotype, thereby providing targets for selection (Furuta et al. 2014).

Our evidence for the involvement of R-M systems in genome rearrangement in intracellular *Endomicrobia* provides a new perspective for future studies targeting the genome evolution of endosymbionts, the evolutionary roles of R-M systems (Rocha et al. 2001) and their potential application as an evolutionary genome engineering tool (Asakura et al. 2011).

Acknowledgements

This work was supported by the Max Planck Society. Hao Zheng is a doctoral student in the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC), Marburg. We thank Lennart Randau and Srivatsa Dwarakanath (MPI, Marburg) for help with discussing the CRISPR-Cas systems.

References

Asakura Y, Kojima H, Kobayashi I. 2011. Evolutionary genome engineering using a restriction-modification system. *Nucleic Acids Res*. 39:9034–9046.

Bennett GM, Moran NA. 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci U S A*. 112:10169–10176.

Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, Weiss RB, Fisher M, Dale C. 2012. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genet*. 8:e1002990.

Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N, Hsu PD, Wu XB, Jiang WY, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 339:819–823.

Dupuis ME, Villion M, Magadan AH, Moineau S. 2013. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat Commun*. 4:2087.

Furuta Y, Abe K, Kobayashi I. 2010. Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Res*. 38:2428–2443.

Furuta Y, Kawai M, Yahara K, Takahashi N, Handa N, Tsuru T, Oshima K, Yoshida M, Azuma T, Hattori M, et al. 2011. Birth and death of genes linked to chromosomal inversion. *Proc Natl Acad Sci U S A*. 108:1501–1506.

Furuta Y, Kobayashi I. 2012. Movement of DNA sequence recognition domains between non-orthologous proteins. *Nucleic Acids Res*. 40:9218–9232.

Furuta Y, Kobayashi I. 2013. Restriction-modification systems as mobile genetic elements. In: Roberts AP ,Mullany P, editors. Bacterial integrative mobile genetic elements. Austin (TX): Landes Bioscience. p. 85–103.

Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, Sugano S, Hasebe M, Kobayashi I. 2014. Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS Genet*. 10:e1004272.

Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P, Brune A. 2009. Genomic analysis of "*Elusimicrobium minutum*," the first cultivated representative of the phylum "*Elusimicrobia*" (formerly Termite Group 1). *Appl Environ Microbiol*. 75:2841–2849.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci U S A*. 105:5555–5560.

Ikeda-Ohtsubo W, Brune A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol Ecol*. 18:332–342.

Kobayashi I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res*. 29:3742–3756.

Li GM. 2008. Mechanisms and functions of DNA mismatch repair. *Cell Res*. 18:85–98.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 10:13–26.

Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet*. 42:165–190.

Moran NA, Bennett GM. 2014. The tiniest tiny genomes. *Annu Rev Microbiol*. 68:195–215.

Nobusato A, Uchiyama I, Ohashi S, Kobayashi I. 2000. Insertion with long target duplication: a mechanism for gene mobility suggested from comparison of two related bacterial genomes. *Gene*. 259:99–108.

Oakeson KF, Gil R, Clayton AL, Dunn DM, von Niederhausern AC, Hamil C, Aoyagi A, Duval B, Baca A, Silva FJ, et al. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol Evol*. 6:76–93.

Ohkuma M, Sato T, Noda S, Ui S, Kudo T, Hongoh Y. 2007. The candidate phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol Ecol*. 60:467–476.

Oliveira PH, Touchon M, Rocha EP. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res*. 42:10618–10631.

Pingoud A. 2004. Restriction endonucleases. Berlin, Heidelberg: Springer.

Rocha EP, Danchin A, Viari A. 1999. Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes. *Mol Biol Evol*. 16:1219–1230.

Rocha EP, Danchin A, Viari A. 2001. Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Res*. 11:946–958.

Sadykov M, Asami Y, Niki H, Handa N, Itaya M, Tanokura M, Kobayashi I. 2003. Multiplication of a restriction-modification gene complex. *Mol Microbiol*. 48:417–427.

Stingl U, Radek R, Yang H, Brune A. 2005. "*Endomicrobia*": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol*. 71:1473–1479.

Takahashi N, Ohashi S, Sadykov MR, Mizutani-Ui Y, Kobayashi I. 2011. IS-Linked movement of a restriction-modification system. *PLoS ONE*. 6:e16554.

Tsuru T, Kawai M, Mizutani-Ui Y, Uchiyama I, Kobayashi I. 2006. Evolution of paralogous genes: reconstruction of genome rearrangements through comparison of multiple genomes within *Staphylococcus aureus*. *Mol Biol Evol*. 23:1269–1285.

Yamaoka I, Nagatani Y. 1977. Cellulose digestion system in the termite, *Reticulitermes speratus* (Kolbe). II. Ultra-structural changes related to the ingestion and digestion of cellulose by the flagellate, *Trichonympha agilis*. *Zool Mag.* 86:34–42.

Zheng H, Dietrich C, Thompson CL, Meuser K, Brune A. 2015a. Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microbes Environ*. 30:92–98.

Zheng H, Dietrich C, Radek R, Brune A. 2015b. *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. *Environ Microbiol*. doi: 10.1111/1462-2920.12960.

Zheng H, Brune A. 2015. Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum *Elusimicrobia*). *Genome Announc*. 3:e00679-15.

Supplementary Information

Restriction-modification systems as mobile genetic elements in the evolution of an intracellular symbiont

Hao Zheng, Carsten Dietrich, Yuichi Hongoh, Andreas Brune

Contents

Materials and Methods

Table S5.1. List of restriction endonuclease and methyltransferase genes in the genome of '*Ca.*Endomicrobium trichonymphae' strain Rs-D17. See Table_S1_RM_RsD17.xlsx (https://www.dropbox.com/s/q8nvjzql5up2mcm/Table_S1_RM_list.xlsx?dl=0)

Table S5.2. List of restriction endonuclease and methyltransferase genes in the genome of *E. proavitum*. See Table_S2_RM_Epr.xlsx (https://www.dropbox.com/s/8tbw4lgszb7qx4q/Table_S2_CRISPR_spacer.xlsx?dl=0)

Table S5.3. The spacer sequences of two CRISPR loci (CR1 and CR2) in the genomes of '*Ca.* Endomicrobium trichonymphae' strain Rs-D17 and *E. proavitum*. See Table_S3_CRISPR_spacer.xlsx

Fig. S5.1. Whole-genome alignment of '*Ca.* Endomicrobium trichonymphae' strain Rs-D17 and *E. proavitum* performed with *Mauve* software.

Fig. S5.2. Cumulative GC skew diagrams of the genomes of '*Ca.* Endomicrobium trichonymphae' strain Rs-D17 and *E. proavitum.*

Fig. S5.3. Different genomic signatures of the R-M genes in the genome of '*Ca.* Endomicrobium trichonymphae' strain Rs-D17.

Fig. S5.4. The CRISPR-Cas systems of '*Ca.* Endomicrobium trichonymphae' strain Rs-D17 and *E. proavitum*.

Fig. S5.5. Repeat sequences of two CRISPR loci (CR1 and CR2) in the genomes of '*Ca.* Endomicrobium trichonymphae' strain Rs-D17 and *E. proavitum*.

Fig. S5.6. Genome rearrangements near the position of all restriction-modification systems and solitary methyltransferase genes in the genome of '*Ca.* Endomicrobium trichonymphae' strain Rs-D17.

Supplementary figures

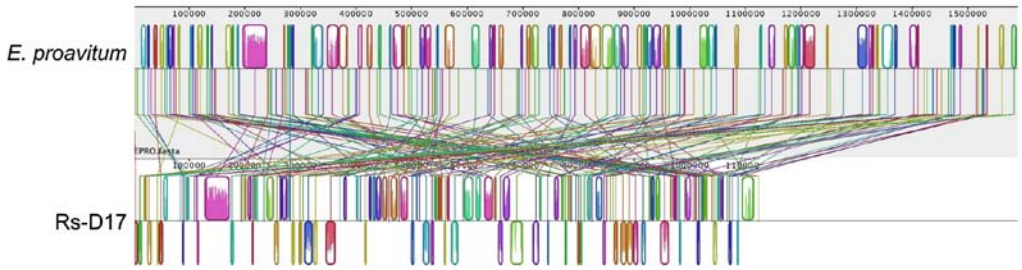

Fig. S5.1. Whole-genome alignment of '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17 and *Endomicrobium proavitum* performed with Mauve software (Darling et al. 2004) showing 166 rearrangenmet events.

Reference

Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394–1403.

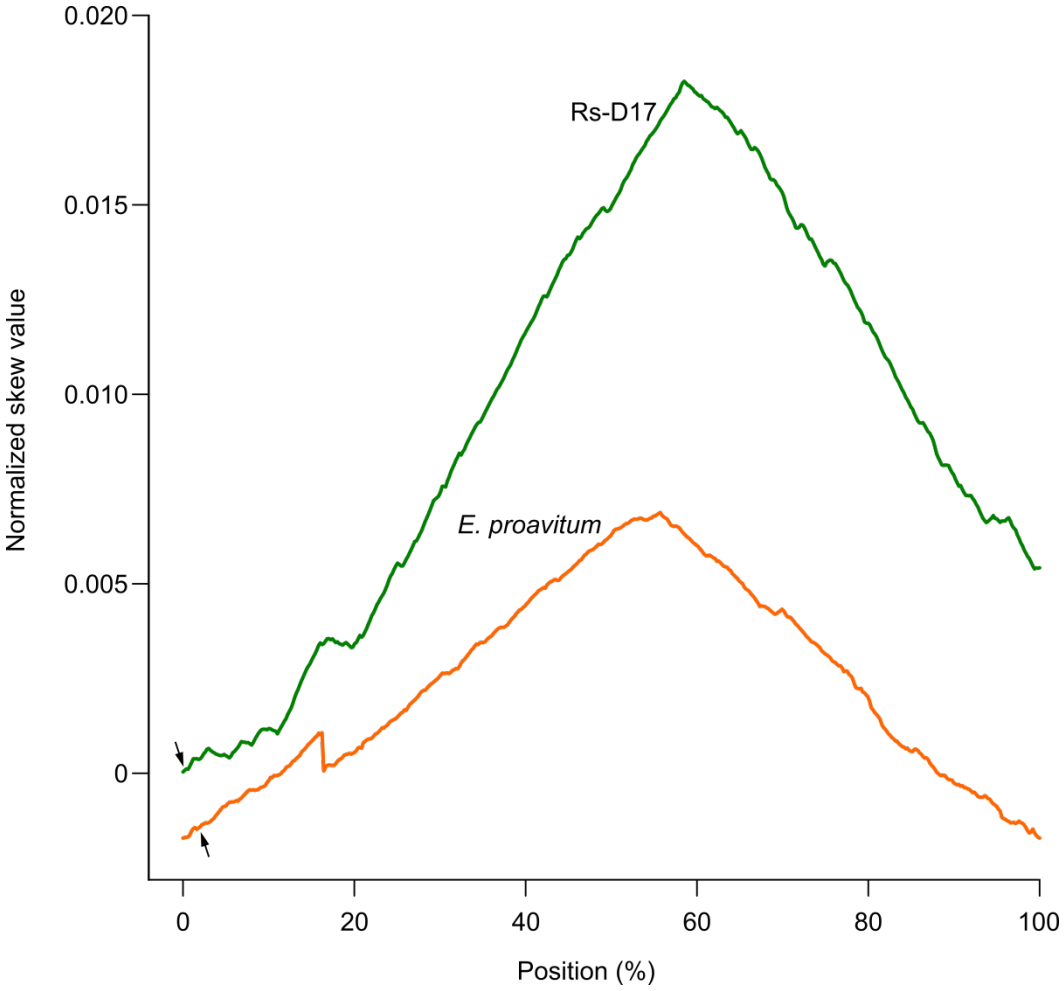

Fig. S5.2. Cumulative GC skew diagrams of the genomes of '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17 and *Endomicrobium proavitum.*

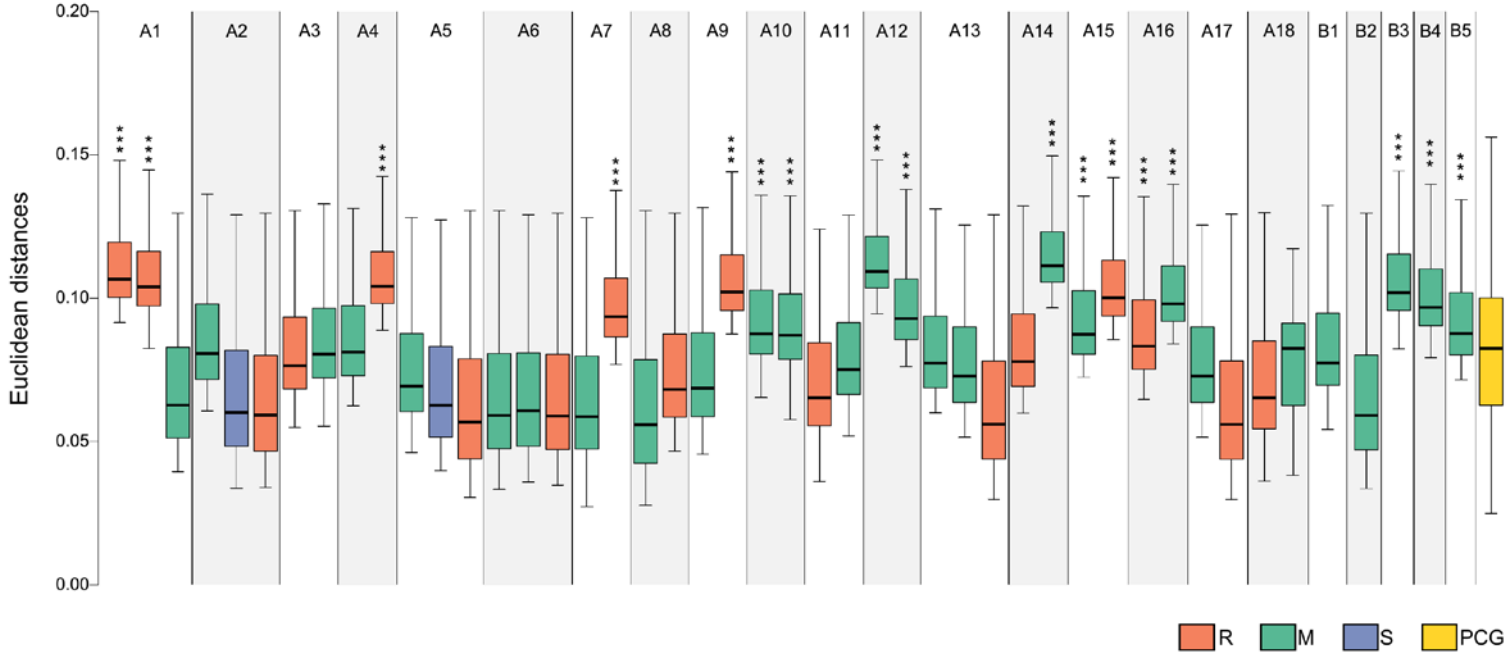

Fig. S5.3. Different genomic signatures of the R-M genes in the genome of '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17. Boxplots show the Euclidean distances between the pentanucleotide frequencies of restriction endonucleases (R), methyltransferases (M), DNA sequence specificity (S) genes and the other intact protein-coding genes (see supplementary Materials and Methods). Genes labeled with *** were significantly ($p < 0.01$) higher than the intragenomic distances of among all protein-coding genes (PCG). Error bars represent standard deviations. The number of R-M genes are as listed in Table S5.1.

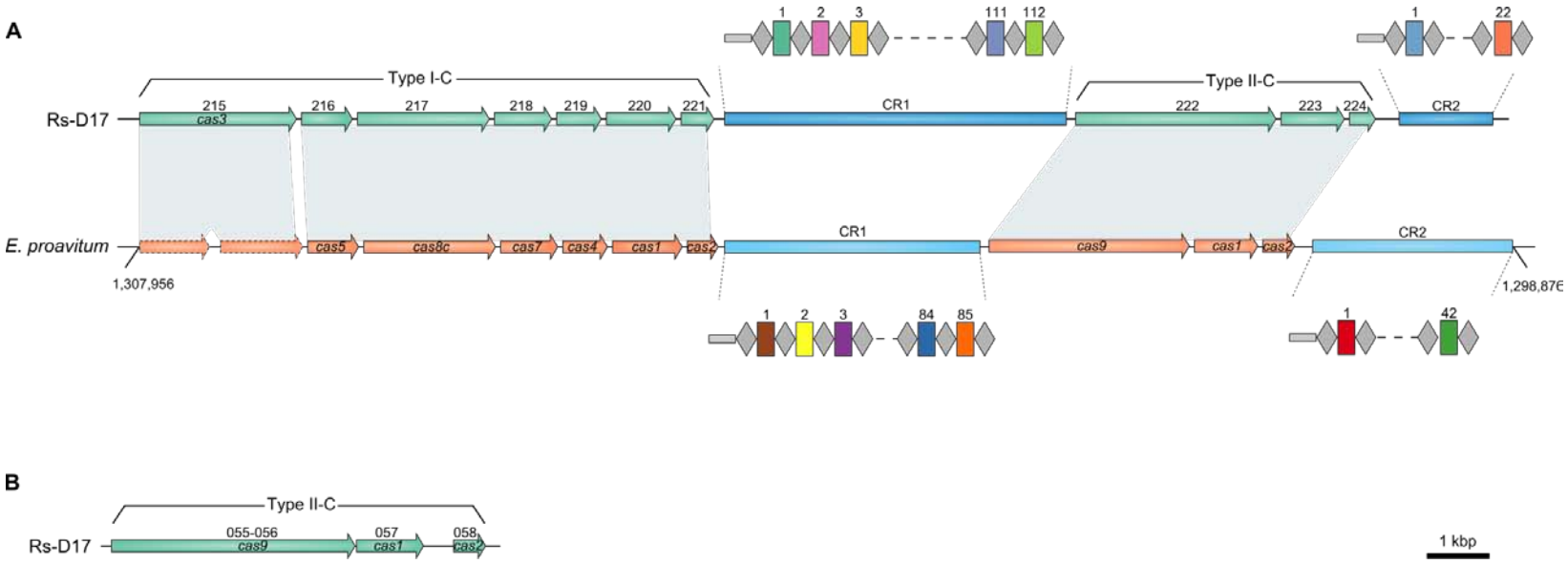

Fig. S5.4. The CRISPR-Cas systems of '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17 and *Endomicrobium proavitum*. (A) Type I-C and type II-C CRISPR-Cas systems consist of 10 *cas* genes and two CRISPR loci (CR1 and CR2). One *cas3* gene is pseudogenized in *E. proavitum* (dashed arrows). For each CRISPR locus, the repeat and spacer content is shown as gray diamonds and colored rectangles, respectively. An AT-rich leader sequence precedes each CRISPR locus (gray rectangle). (B) A cluster of genes of the type II-C CRISPR-Cas system without a CRISPR locus in the genome of strain Rs-D17. The sequences of the CR1 and CR2 consensus repeats are shown in Fig. S5.3 and each spacer is given in Table S5.2. All genes follow the nomenclature of Makarova et al. (2011).

References

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, et al. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*. 9:467-477.

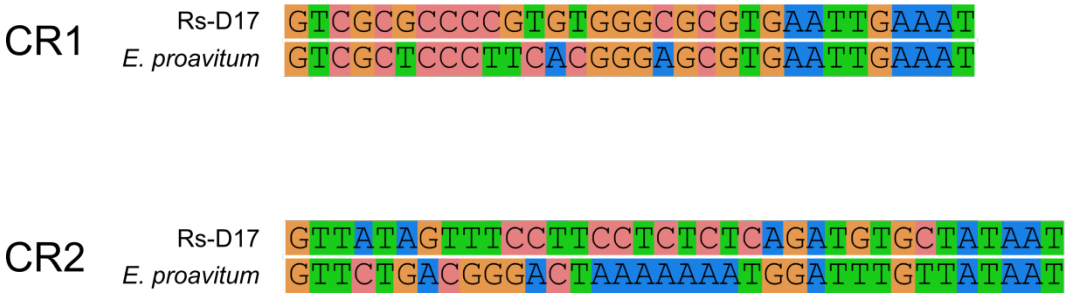

Fig. S5.5. Repeat sequences of two CRISPR loci (CR1 and CR2) in the genomes of '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17 and *Endomicrobium proavitum*. The length of the repeat sequences in the two loci are the same in the two organisms: 32 bp (CR1) and 36 bp (CR2), respectively. The sequences of CR1 are similar, while the sequences of CR2 are more diverse. The alignment was visualized using SeaView (Gouy et al. 2010).

References

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 27:221-224.

Fig. S5.6. Genome rearrangements near the position of all restriction-modification (RM) systems and solitary methylase genes in the genome of '*Candidatus* Endomicrobium trichonymphae' strain Rs-D17. All RM systems and solitary methylase genes are numbered accordingly as listed in Table S5.1 BA1–A16; B1–B6. Different colors for genes are described in Fig. 5.3.

(https://www.dropbox.com/s/ey7y99f1otr1ga5/SI_Zheng_et_al_RM_mobile_elements.pdf?dl=0)

Chapter 6

General Discussion

The presence of intracellular bacteria in the cytoplasm of flagellates has been observed already in numerous microscopic studies (e.g., Stingl et al. 2015; Sato et al. 2009). Thanks to culture-independent techniques, many termite-specific lineage of microorganisms have been identified, specifically *Endomicrobia* (phylum *Elusimicrobia*). Since the discovery of this new phylum of bacteria, it has attracted numerous studies on its diversity, distribution, *in situ* localization, evolutionary history, and functional roles (reviewed in Brune 2012).

This dissertation described the isolation of *Endomicrobia proavitum*, the first cultured representative of the *Endomicrobia* class and a close relative of '*Candidatus* Endomicrobium trichonymphae' colonizing flagellates of the same termite species. This provides a perfect model for the studies on genome evolution mechanisms during the transition lifestyle from free-living to intracellular symbionts.

The first isolate of the class *Endomicrobia*

The first cultivated representative of the phylum *Elusimicrobia*, *E. minutum*, was obtained from the sterile-filtered gut homogenate from a scarab beetle larva fortuitously (Geissinger et al. 2009; Brune 2014). Inspired by the small size of the endosymbionts, I tried to physically enrich the free-living *Endomicrobia* from filtrated termite gut homogenate and this strategy enabled the isolation of *E. proavitum* (Chapter 3). The fact that the two isolates representing separate distant classes of the phylum both show small cell sizes, which may indicate that the definition of ultramicrobacterium applies to the entire phylum with the cell volumes from 0.007 to 0.177 μm^3 (Chapter 3). Considering they all possess relatively small genomes (1.13–1.64 Mbp; Chapter 5; Appendix), it fits the logic that a smaller cell volume contains a smaller amount of DNA to give enough

space for other cellular components, as suggested for the marine oligotrophs (Dufresne et al. 2003). Interestingly, both *Endomicrobium* strains form tube-like elongations at the tapered cell poles (Chapter 3), which were not observed in *E. minutum*. In addition, *E. proavitum* shows an abnormal coccoid swelling at poles of the rods, which suggests a budding reproduction. Similar pleomorphism of bacterial cells, with both rods and budding cells has also been observed in *Legionella pneumophila*, an aerobic intracellular pathogen (Katz et al. 1984). Both *E. proavitum* and *E. minutum* show abnormall cell cycles, this could be also a common property of the phylum. The reason and mechanism for this unusual cell form still needs to be studied.

The genome of *E. proavitum* carries only one copy of the Group IV nitrogenase operon, and this cluster of nitrogenases have not yet been characterized (Staples et al. 2007). Unlike other bacteria that contain several paralogous nitrogenase gene copies within their genomes (Raymond et al. 2004), *E. proavitum* has only one copy of the *nif* operon in the genome, which enabled us to define its nitrogen-fixing ability without being confounded. The firmly documented N_2-fixing ability of *E. proavitum* by its N_2-dependent growth, $^{15}N_2$ incorporation, and acetylene reduction activity, provides the first evidence that Group IV nitrogenase has the N_2-reducing potential. This result extends our knowledge about phyla of bacteria with diazotrophic representatives and the N_2-fixing ability of the uncharacterized group of nitrogenase. Since the Our finding sheds new light on biology, biochemistry, ecology and evolution of nitrogen fixation. The pure culture of *E. proavitum* offers the opportunity for the further biochemical examination of this group of nitrogenase.

Strictly vertical transmission of *Endomicrobia*

Ikeda-Ohtsubo and Brune (2009) showed a congruent topology of the SSU rRNA gene trees of the endosymbionts and their flagellate hosts, but it was still not excluded that the presence of small populations of a second *Endomicrobia* phylotype within one *Trichonympha* cell. With a highly resolved analysis of the ITS regions, which is more sufficient for highly similar sequences, we investigated the community structure in micromanipulated individual host cells. The fact that each *Trichonympha* cell harbors a homogeneous population of *Endomicrobia* symbionts suggests that the *Endomicrobia* endosymbionts do not only show congruent phylogeny with the flagellates, but also are strictly vertically transmitted by the respective host species, which is also supported that the *Trichonympha* flagellates lost the sexual reproduction ability compared to those in the ancestral cockroach.

Strictly vertical transmission through host lineages has a crucial effect on endosymbiont population structure. This specific transmission mode might impose a population bottleneck on the number of bacteria that are passed from mother to offspring (Hosokawa et al. 2006). The homogenous population of the endosymbionts that are specific for respective host species indicates that *Endomicrobia* endosymbionts also suffered a frequent population bottleneck during transmission (Chapter 2), as other endosymbionts in insects transmitted by bacteriocytes (Koga et al. 2012). These successive bottlenecks throughout the evolution of the bacterial partner reduce their effective population size, which elevates the rate of fixation of deleterious mutations through genetic drift (Wernegreen 2002). Numerous studies implicate the small population size and asexuality of the relevant species, combined with an inherent deletional bias in bacteria, are the major causes of the genome erosion of endosymbionts (Mira et al.

2001). Accordingly, the genome of '*Ca*. Endomicrobium trichonymphae' is slightly smaller than its ancestral free-living relative (Chapter 5) and showed an accumulation of pseudogenes (Hongoh et al. 2008a), which indicates that the *Endomicrobia* endosymbionts are on an early stage of the genome reduction.

Genome reduction of *Endomicrobia*

Although the genome of '*Ca*. Endomicrobium trichonymphae' shows massive evidence of decay, the lack of genomic information of a free-living relative hampered the detailed study of the genome evolution of this endosymbiosis. The pure culture and the complete genome sequence of a close and free-living ancestral relative (Chapter 3; Appendix) provides a perfect template for the comparative analyses. It has been shown that the gene repertoires of an endosymbiont in other insects typically are only a reduced subset from the free-living relative (Santos-Garcia et al. 2014). However, strain Rs-D17 seems to have acquired new genes after the divergence of symbiotic lineages, such as glucose-6-phosphate (G6P) transporter and alcohol/acetaldehyde dehydrogenase (Chapter 5). The newly up-taken UhpC transporter enables the endosymbionts utilize G6P as energy and carbon source in the cytoplasm of flagellate, where the free-sugar must be extremely limited. The Adh alters the fermentation strategy of strain Rs-D17 to recover NAD^+ through the ethanol fermentation. The genome of the symbionts is not simply a subset of its free-living relative, and the gain of new genes might help the process of this intracellular symbiosis. The endosymbionts share the synteny in many pathways, such as amino acid and cofactor synthesis. The ability to synthesize amino acids is also supported by the physiological data (Chapter 3).

Considering the relatively young age of the symbiosis of *Endomicrobia* and their hosts, which is dated to 40-70 Mya (Ikeda-Ohtsubo and Brune 2009), they

are on an early process of the transition, compared to some ancient symbionts (Moran et al. 2008). It is normal to see a purifying selection during this stage, in the case of *Endomicrobia*, we show that this selection force is genome-wise even and does not show bias based on the functions (Chapter 5). It has been shown that genome reduction might occur in several stages, depending on the age of the association (McCutcheon and Moran 2012). In more anciently evolved symbionts, genomes represent the most reduced stages of a progression of genomic changes that result from the small population size and asexuality (Shigenobu et al. 2000). The proliferation of mobile genetic elements (MGEs) and multiple genomic rearrangements are usually observed in the early stage of genome reduction (Koga and Moran 2014; Manzano-Marín and Latorre 2014; Oakeson et al. 2014). These MGEs have also facilitated numerous inactivation and deletion in various genes in connection with genome rearrangements (Burke and Moran 2011; Manzano-Marín and Latorre 2014; Oakeson et al. 2014). To date, only conventional mobile elements, such as transposons and phage-associated genes, have been implicated in this process. This thesis documents the first case where restriction-modification (RM) systems acting as MGEs are responsible for the genome rearrangements (Chapter 4). This will help us to better understand the mechanisms of genome reduction during the initial stages of endosymbiont evolution. It was shown that flagellates from various genera all possess *Endomicrobia* endosymbionts, which belong to diverse lineages (Chapter 1; Chapter 2). Thus far, only one complete genome of the endosymbiont is available, more genomic information of the endosymbionts from different lineage will provide more knowledge of such endosymbiosis.

The genome reduction of *Endomicrobia* intracellular symbionts show several features: (i) Although many pathways are disrupted by pseudogenes or gene

losses, they have also acquired several genes to assist the symbiosis process; (ii) The synthesis pathways of outer membrane is collapsed as in other endosymbionts; (iii) massive genome rearrangements occurred; (iv) RM systems act as MGEs in the genome rearrangements, instead of other conventional mobile elements.

Flagellate and bacterial symbionts

Recently, a highly resolved phylogeny of the major *Endomicrobia* lineages using concatenated 16S and 23S rRNA gene sequences revealed the presence of distinct lineages of free-living and endosymbiotic lineages in a wide range of dictyopteran host (Thompson et al., unpublished data). This resulted indicated suggest that different free-living lineages of *Endomicrobia* were acquired as flagellate endosymbionts over the course of termite evolution. Additional genome sequences of both endosymbiotic and free-living *Endomicrobia* will provide a better understand of the mechanism of the reductive evolution that occurred upon association with the flagellate hosts, which happend several times during the evolution.

The cellulolytic flagellated protists are consistently associated with prokaryotic cells. They colonize both the surface and cytoplasm and sometimes even the nucleus of their hosts cells (Brune and Dietrich 2015). Phylogenetic analyses revealed the co-speciation between such symbionts and the flagellate hosts; accordingly, it is believed that the symbionts confer beneficial traits to the hosts. Because the endosymbionts are mainly host-dependent, it is not possible to get pure culture of the symbionts. Thus far, several studies have documented the genome sequences of flagellate symbionts from different phyla of bacteria and predicted their potential functional roles (Table 6.1). Considering that there are various associations with termite gut protists involving *Endomicrobia*,

Spirochaetes, *Bacteroidetes*, and methanogens, and given that a great diversity of flagellates still remain to be examined with respect to molecular identification of associated bacteria, the termite gut will provide rich and attractive models for studying cellular symbioses.

Table 6.1. Flagellate symbionts in the termite gut with genomic information.

	Size (Mbp)	G + C Content (%)	Flagellate host	Termite host	Reference
'*Ca.* Endomicrobium trichonymphae'	1.13	35	*T. agilis*	*R. speratus*	(Hongoh et al. 2008a)
'*Ca.* Azobacteroides pseudotrichonymphae'	1.11	33	*P. grassii*	*C. formosanus*	(Hongoh et al. 2008b)
'*Ca.* Treponema intracellularis'	2.7–3.2	55	*Eucomonympha* sp.	*H. sjoestedti*	(Ohkuma et al. 2015)
'*Ca.* Adiutrix intracellularis'	2.3	43	*T. collaris*	*Z. nevadensis*	(Ikeda-Ohtsubo et al., submitted)
'*Ca.* Ancillula trichonymphae'	1.48	44	*T. paraspiralis*	*I. marginipennis*	(Strassert et al., submitted)
'*Ca.* Symbiothrix dinenymphae'	4.3	46	*Dinenympha* sp.	*R. speratus*	(Yuki et al. 2015)

References

Brune A. 2012. Endomicrobia: intracellular symbionts of termite gut flagellates. *J Endocytobiosis Cell Res*. 23:11–15.

Brune A. 2014. The family *Elusimicrobiaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E ,Thompson F, editors. The Prokaryotes. Berlin: Springer, pp. 637–640.

Brune A, Dietrich C. 2015. The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. *Annu Rev Microbiol*. doi: 10.1146/annurev-micro-092412-155715.

Burke GR, Moran NA. 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol*. 3:195–208.

Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V, Duprat S, Galperin MY, Koonin EV, Le Gall F, et al. 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci U S A*. 100:10020–10025.

Geissinger O, Herlemann DPR, Möschel E, Maier UG, Brune A. 2009. The ultramicrobacterium "*Elusimicrobium minutum*" gen. nov., sp. nov., the first cultivated representative of the Termite Group 1 phylum. *Appl Environ Microbiol*. 75:2831–2840.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008a. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci U S A*. 105:5555–5560.

Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, et al. 2008b. Genome of an endosymbiont coupling N_2 fixation to cellulolysis within protist cells in termite gut. *Science*. 322:1108–1109.

Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T. 2006. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biol*. 4:1841–1851.

Ikeda-Ohtsubo W, Brune A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol Ecol*. 18:332–342.

Katz SM, Hashemi S, Brown KR, Habib WA, Hammel JM. 1984. Pleomorphism of *Legionella pneumophila*. *Ultrastruct Pathol*. 6:117–129.

Koga R, Meng XY, Tsuchida T, Fukatsu T. 2012. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proc Natl Acad Sci U S A*. 109:E1230–E1237.

Koga R, Moran NA. 2014. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *ISME J*. 8:1237–1246.

Manzano-Marín A, Latorre A. 2014. Settling down: the genome of *Serratia symbiotica* from the aphid *Cinara tujafilina* zooms in on the process of accommodation to a cooperative intracellular life. *Genome Biol Evol*. 6:1683–1698.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 10:13–26.

Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet*. 17:589–596.

Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet*. 42:165–190.

Oakeson KF, Gil R, Clayton AL, Dunn DM, von Niederhausern AC, Hamil C, Aoyagi A, Duval B, Baca A, Silva FJ, et al. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome BiolEvol*. 6:76–93.

Ohkuma M, Noda S, Hattori S, Iida T, Yuki M, Starns D, Inoue J, Darby AC, Hongoh Y. 2015. Acetogenesis from H_2 plus CO_2 and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. *Proc Natl Acad Sci U S A*. 112:10224–10230.

Raymond J, Siefert JL, Staples CR, Blankenship RE. 2004. The natural history of nitrogen fixation. *Mol Biol Evol*. 21:541–554.

Santos-Garcia D, Latorre A, Moya A, Gibbs G, Hartung V, Dettner K, Kuechler SM, Silva FJ. 2014. Small but powerful, the primary endosymbiont of moss

bugs, *Candidatus* Evansia muelleri, holds a reduced genome with large biosynthetic capabilities. *Genome Biol Evol*. 6:1875–1893.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*. 407:81–86.

Staples CR, Lahiri S, Raymond J, Von Herbulis L, Mukhophadhyay B, Blankenship RE. 2007. Expression and association of group IV nitrogenase NifD and NifH homologs in the non-nitrogen-fixing archaeon *Methanocaldococcus jannaschii*. *J Bacteriol*. 189:7392–7398.

Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet*. 3:850–861.

Yuki M, Kuwahara H, Shintani M, Izawa K, Sato T, Starns D, Hongoh Y, Ohkuma M. 2015. Dominant ectosymbiotic bacteria of cellulolytic protists in the termite gut also have the potential to digest lignocellulose. *Environ Microbiol*. doi: 10.1111/1462-2920.12945.

Appendix

Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum *Elusimicrobia*)

Published in *Genome Announcement* (2015) 3:e00679-15.

Hao Zheng[1], Andreas Brune[1,2]

[1]Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany; [2]LOEWE Center for Synthetic Microbiology (SYNMIKRO), Philipps-Universität Marburg, Marburg, Germany

Contributions: H.Z. planned and carried out experiments, analyzed data, discussed results, and wrote the manuscript. A.B. discussed results, wrote the manuscript, and secured funding.

Abstract

We sequenced the complete genome of *Endomicrobium proavitum* strain Rsa215, the first isolate of the class *Endomicrobia* (phylum *Elusimicrobia*). It is the closest free-living relative of the endosymbionts of termite gut flagellates, and thereby provides an excellent model for studying the evolutionary processes during the establishment of an intracellular symbiosis.

Genome announcement

Endomicrobia (formerly Termite Group 1) form a deep-branching lineage of uncultivated bacteria in the *Elusimicrobia* phylum (Brune 2012, 2014), which is so far represented only by a single isolate, *Elusimicrobium minutum* (Geissinger et al. 2009). Members of this clade have been identified as intracellular symbionts of termite gut flagellates and are specific for and vertically transferred by their respective hosts (Stingl et al. 2005; Ikeda-Ohtsubo and Brune 2009; Zheng et al. 2015a). However, the recovery of 16S rRNA genes of Endomicrobia from artificially defaunated or flagellate-free termites and cockroaches indicated the existence of putatively free-living relatives in the same habitat (Ohkuma et al. 2007; Ikeda-Ohtsubo et al. 2010). We isolated the strictly anaerobic ultramicrobacterium *Endomicrobium proavitum* strain Rsa215, the first representative of the class *Endomicrobia*, from sterile-filtered gut homogenate of *Reticulitermes santonensis* (Zheng et al. 2015b).

The genome sequence of *E. proavitum* was obtained with Pacific Biosciences (PacBio) SMRT sequencing. Genomic DNA of strain Rsa215 was prepared using cetyltrimethylammonium bromide (CTAB) extraction (Winnepenninckx et al. 1993) and commercially sequenced (GATC Biotech AG, Germany) on a PacBio RS platform using three SMRT cells (insert size 8–12 kbp). Reads were assembled using PacBio's SMRT Portal software (v2.1.0) and the HGAP assembly algorithm (Chin et al. 2013). Initial assembly yielded seven contigs with more than 20× average coverage. Reassembly with Minimus 2 (Treangen et al. 2011) identified overlaps of all contigs, which were confirmed by PCR and Sanger sequencing experiments. The final assembly resulted in a circular chromosome (1,588,979 bp, 39 mol% G + C content) and was confirmed using the latest version of the assembler (HGAP 3). Plasmids were not detected.

Annotation on the *MicroScope* platform (Vallenet et al. 2013) resulted in 1,345 predicted protein-coding genes, 46 tRNA genes, and a single set (16S, 23S, and 5S) of rRNA genes. Many genes of *E. proavitum* were highly similar to those of the closely related "*Candidatus* Endomicrobium trichonymphae" strain Rs-D17 (Hongoh et al. 2008), which is an endosymbiont of the large flagellates that co-occur in the same habitat (Stingl et al., 2005). Although the two organisms share many metabolic pathways, the genome of the endosymbiont is much smaller (1,125,857 bp) and highly degraded, and numerous pathways (e.g., initiation and regulation of chromosomal replication, lipopolysaccharide biosynthesis, ammonium transport, and assimilation) are interrupted by pseudogenes (Hongoh et al. 2008). The same pathways are intact in the free-living *E. proavitum*, which shows no obvious signs of genome reduction but possesses additional functions absent from the endosymbiont, including a set of genes required for nitrogen fixation (*nifHDK*). Comparative genome analysis of the two strains will provide a better understanding of the evolutionary processes that started when their common free-living ancestor became associated with its flagellate host.

Nucleotide sequence accession number

The sequence data for the genome have been deposited in DDBJ/EMBL/Genbank under the accession no. CP009498.

Acknowledgments

This study was funded by the Max Planck Society. Hao Zheng is a doctoral student in the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC), Marburg.

References

Brune A. 2012. Endomicrobia: intracellular symbionts of termite gut flagellates. *J Endocytobiosis Cell Res*. 23:11–15.

Brune A. 2014. The family *Elusimicrobiaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E ,Thompson F, editors. The Prokaryotes. Berlin: Springer, pp. 637–640.

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 10:563–569.

Geissinger O, Herlemann DPR, Möschel E, Maier UG, Brune A. 2009. The ultramicrobacterium "*Elusimicrobium minutum*" gen. nov., sp. nov., the first cultivated representative of the Termite Group 1 phylum. *Appl Environ Microbiol*. 75:2831–2840.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci U S A*. 105:5555–5560.

Ikeda-Ohtsubo W, Brune A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol Ecol*. 18:332–342.

Ikeda-Ohtsubo W, Faivre N, Brune A. 2010. Putatively free-living '*Endomicrobia*' – ancestors of the intracellular symbionts of termite gut flagellates? *Environ Microbiol Rep*. 2:554–559.

Ohkuma M, Sato T, Noda S, Ui S, Kudo T, Hongoh Y. 2007. The candidate phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol Ecol*. 60:467–476.

Stingl U, Radek R, Yang H, Brune A. 2005. "*Endomicrobia*": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol*. 71:1473–1479.

Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. 2011. Next generation sequence assembly with AMOS. *Curr Protoc Bioinformatics*. 33:11.18.11–11.18.18.

Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fevre F, Longin C, Mornico D, Roche D, et al. 2013. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res*. 41:D636–D647.

Winnepenninckx B, Backeljau T, De Wachter R. 1993. Extraction of high molecular weight DNA from molluscs. *Trends Genet*. 9:407.

Zheng H, Dietrich C, Thompson CL, Meuser K, Brune A. 2015a. Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microbes Environ*. 30:92–98.

Zheng H, Dietrich C, Radek R, Brune A. 2015b. *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. *Environ Microbiol*. doi: 10.1111/1462-2920.12960.

Acknowledgements

I would like to thank

Prof. Dr. Andreas Brune for supervising me and for continued support and guidance from day one when I arrived Germany. I thank him for spending so much time on training my way of thinking on research, my writing and oral presenting skill. His preciseness on scientific research will affect my whole life;

Prof. Dr. Uwe Maier, Prof. Dr. Martin Thanbichlwe and Prof. Dr. Hans-Ulrich Mösch for agreeing to examine this dissertation and offering their comments;

my thesis committee members, Prof. Dr. Uwe Maier, Prof. Dr. Eva Stukenbrock, PD Dr. Werner Liesack, for their support during these three years.

International Max Planck Research School for Environmental, Cellular and Molecular Microbiology for funding my study;

every member of AG Brune and Department Biogeochemistry for their help in many aspects, especially Katja Meuser and Dr. Claire L. Thompson for their support during the very begging when I join the lab. I am very happy to collaborate with Dr. Carsten Dietrich a lot for my projects.

I sincerely thank my parents and my whole family in China for their selfless love and encouragement. My dear wife Xue Zhang for all her and her family's support. My lovely son Zechuan Zheng for not being so noisy, Dad loves you forever.

Pledge

I certify that the present thesis entitled:

"**Genome evolution of *Endomicrobia*: from free-living bacteria to intracellular symbionts of termite gut flagellates**"

was carried out without any unlawful means. This work has never been submitted before in this or in a similar format to any other university and has not been used before any examination.

Hao Zheng

Marburg, September 2015

Curriculum Vitae

Name	Hao Zheng
Date of birth	15th Feb. 1985
Place of birth	Changchun, Jilin, China
Marital status	Married
Nationality	Chinese

Education

Max Planck Institute for Terrestrial Microbiology Oct. 2012 – present
Doctoral studies

Tokyo Institute of Technology and Tsinghua University Aug. 2009 – Aug. 2012
Dual Degree: Master of Engineering

Tsinghua University, Beijing Aug. 2005 – Aug. 2009
Bachelor of Engineering

Publications

Zheng H, Dietrich C, Hongoh Y, Brune A. 2015. Restriction-modification systems as mobile genetic elements in the evolution of an intracellular symbiont. *Mol Biol Evol*. Doi:10.1093/molbev/msv264.

Zheng H, Dietrich C, Radek R, Brune A. 2015. *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. *Environ Microbiol.* doi:10.1111/1462-2920.12960.

Zheng H, Brune A. 2015. Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum *Elusimicrobia*). *Genome Announc.* 3: e00679-15.

Zheng H, Dietrich C, Thompson CL, Brune A. 2015. Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microbes Environ.* 30:92–98.

Zheng H, Bodington D, Zhang C, Tanji Y, Miyanaga K, Hongoh Y, Xing XH. 2013. Comprehensive phylogenetic diversity of [FeFe]-hydrogenase genes in termite gut microbiota. *Microbes Environ.* 28: 491–494.

Zheng H, Zhang C, Lu Y, Jiang PX, Xing XH. 2012. Alteration of anaerobic metabolism in Escherichia coli for enhanced hydrogen production by heterologous expression of hydrogenase genes originating from the *Synechocystis* sp.. *Biochem Eng J*. 60: 81–86.

Liu Z, Lv FX, **Zheng H**, Zhang C, Wei F, Xing XH. 2012. Enhanced hydrogen production in a UASB reactor by retaining microbial consortium onto carbon nanotubes (CNTs). *Int J Hydrogen Energ.* 37: 10619–10626.

Feng Q, Wang YX, Wang TM, **Zheng H**, Chu LB, Zhang C, Chen HZ, Kong XQ, Xing XH. 2012. Effects of packing rates of cubic-shaped polyurethane foam carriers on the microbial community and the removal of organics and nitrogen in moving bed biofilm reactors. *Bioresource Technol.* 117: 201–207.

Yan ST, **Zheng H**, Li A, Zhang X, Xing XH, Chu LB, Ding GJ, Sun XL, Jurcik B. 2009. Systematic analysis of biochemical performance and the microbial community of an activated sludge process using ozone-treated sludge for sludge reduction. *Bioresource Technol.* 100: 5002–5009.

Yan ST, Li A, **Zheng H**, Luo MF, Xing XH. 2009. Effects of ionic surfactants on bacterial luciferase and α-amylase. *Chinese J Chem Eng.* 17: 829–834.

Honors

Fellowship from the International Max Planck Research School.

Scholarship from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Tsinghua University First-Class Scholarship in 2007, 2010 and 2011.